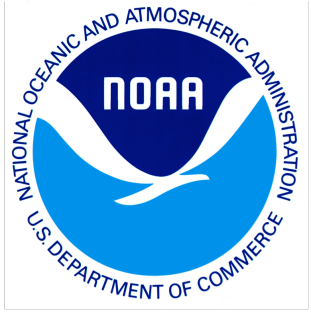
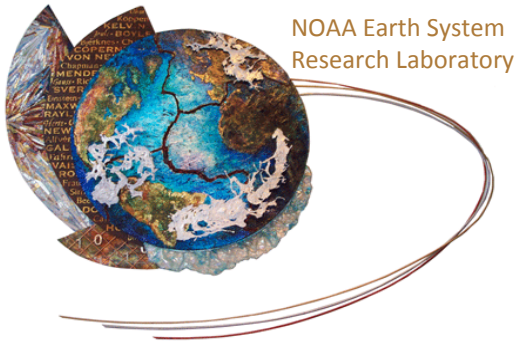




Ensemble prediction and winter weather

Tom Hamill
NOAA Earth System Research Lab
tom.hamill@noaa.gov
(303) 497-3060

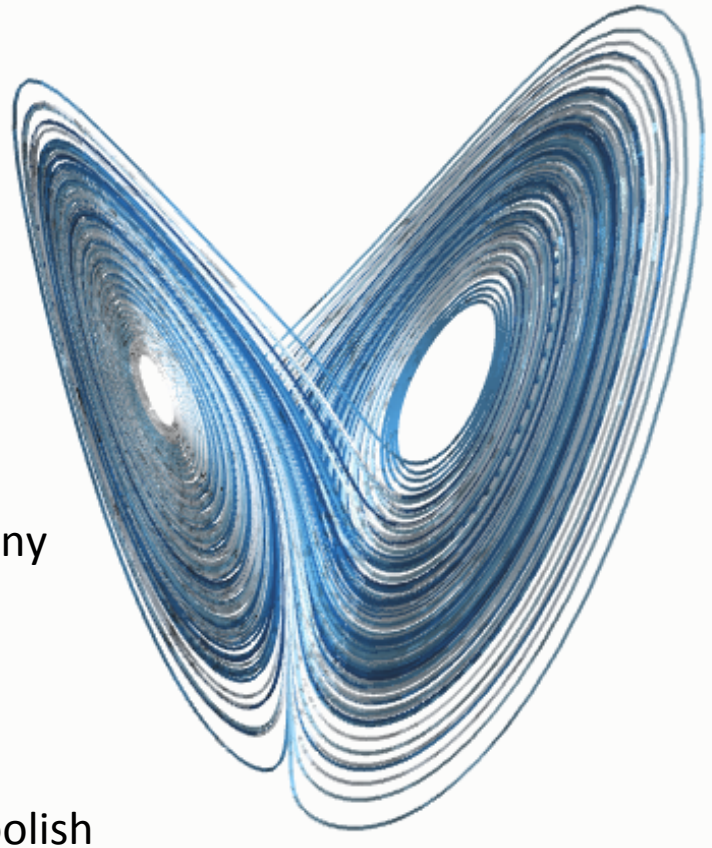
Organization of this talk.

- Part 1: Ensemble theory and nuts and bolts.
  - Motivation for ensembles
  - How we construct them.
  - Advantages/disadvantages of some various ensemble approaches.

- Part 2: Using ensembles for winter weather.

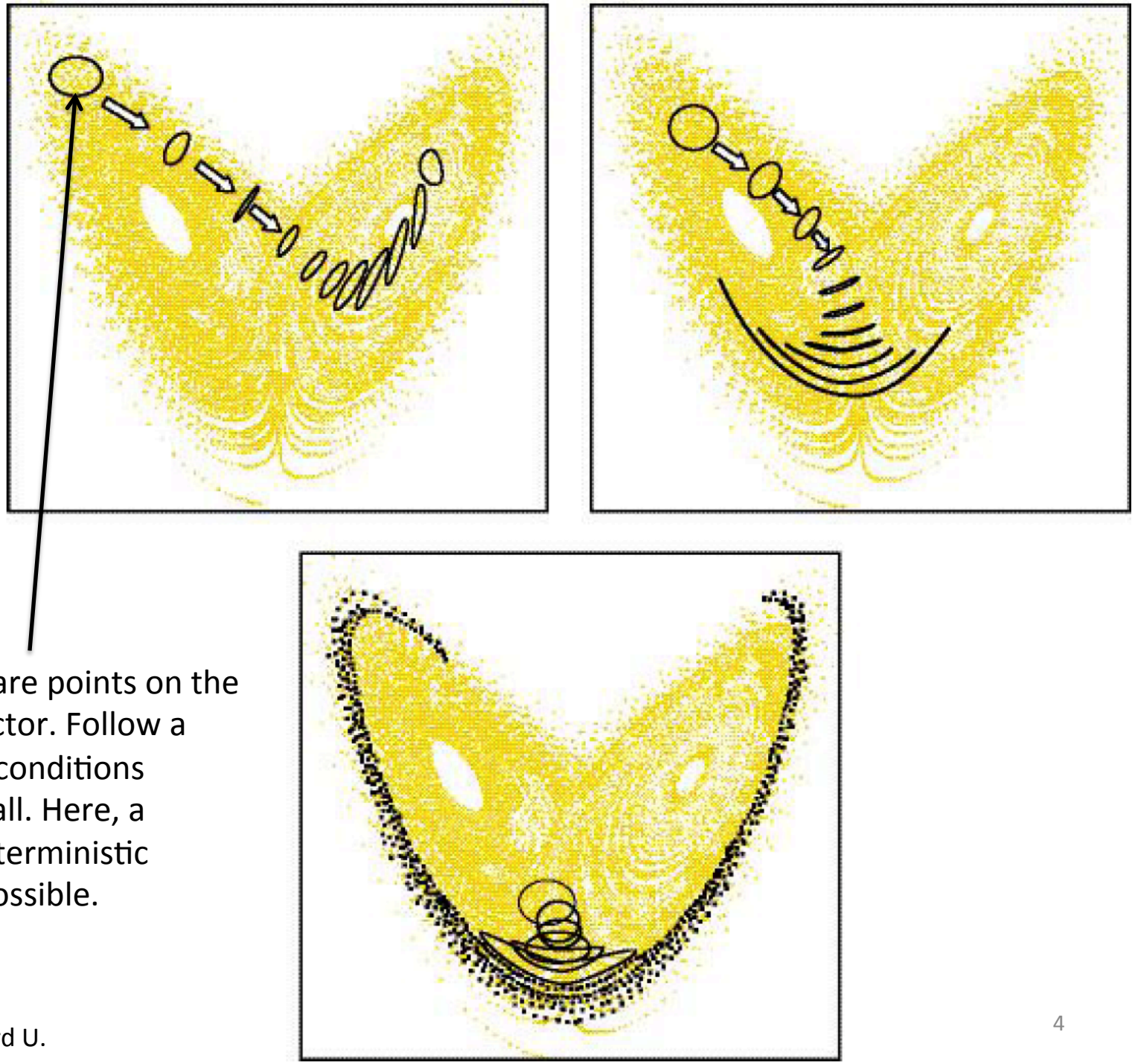
The Lorenz (1963) dynamical system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$



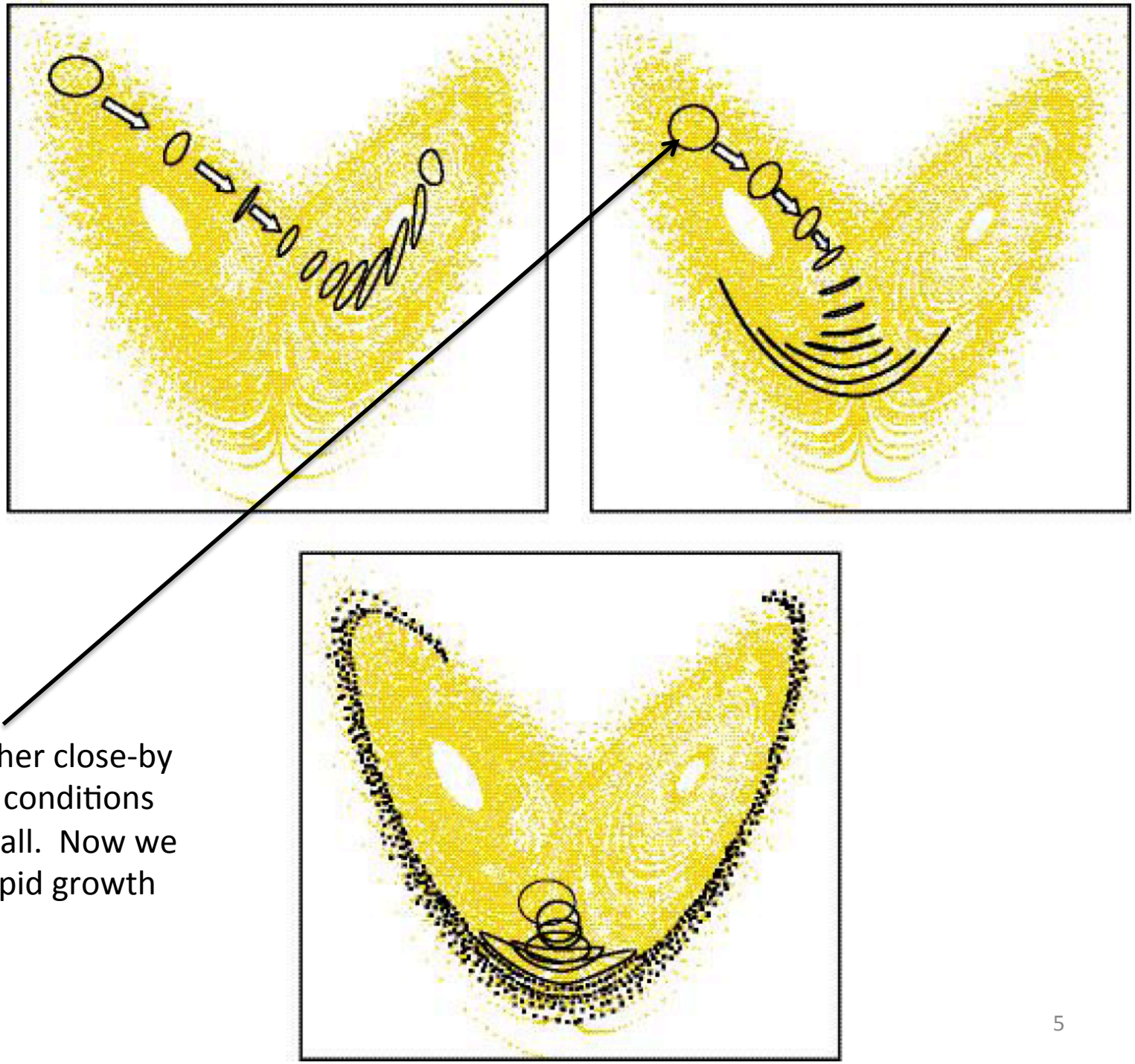
A 3-dimensional dynamical system that illustrates the property of "chaos." Here is a picture of the Lorenz attractor. Start off with any (x,y,z) value, and very quickly the state will begin spiraling around in this reduced set of points.

The atmosphere has an attractor. Were you foolish enough to initialize a weather forecast model with an 80F temperature at the north pole, the forecast state would quickly cool.

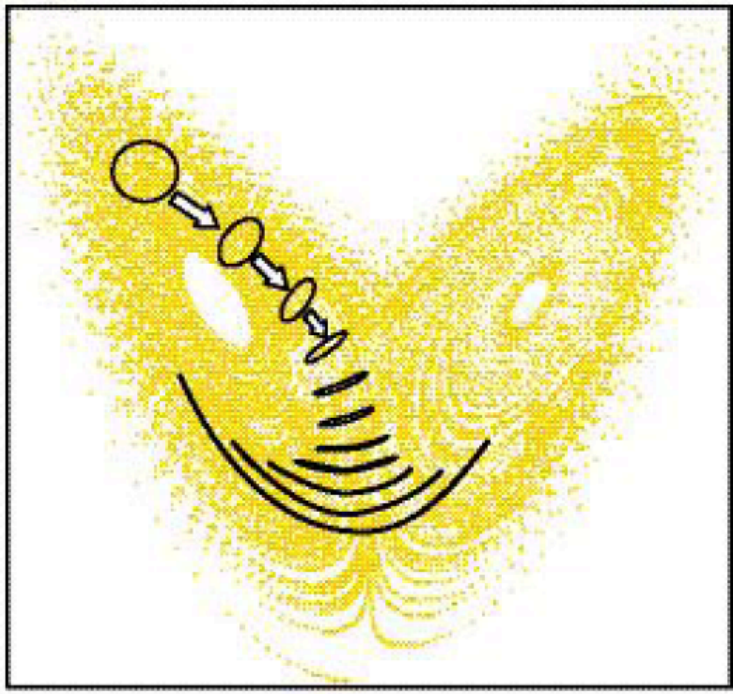


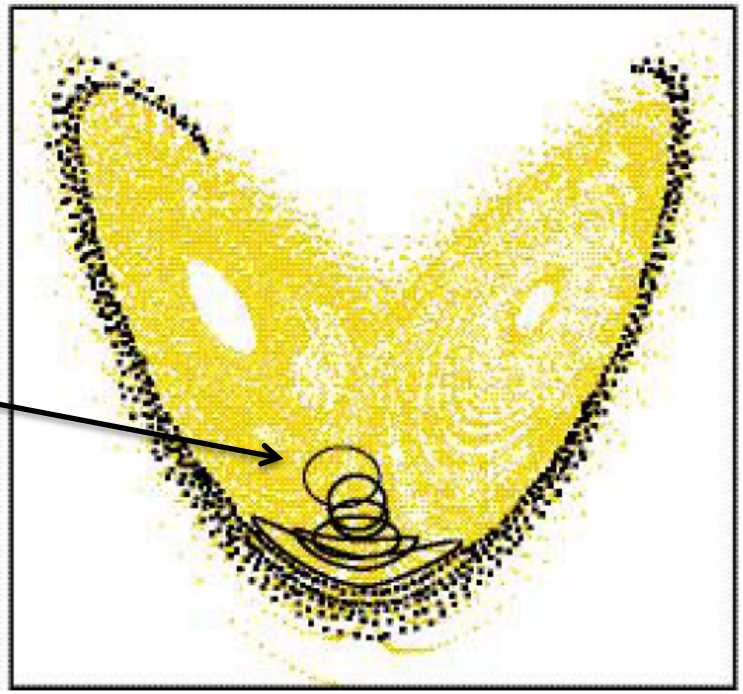
Yellow dots are points on the Lorenz attractor. Follow a set of initial conditions within the ball. Here, a relatively deterministic forecast is possible.

c/o Tim Palmer, Oxford U.



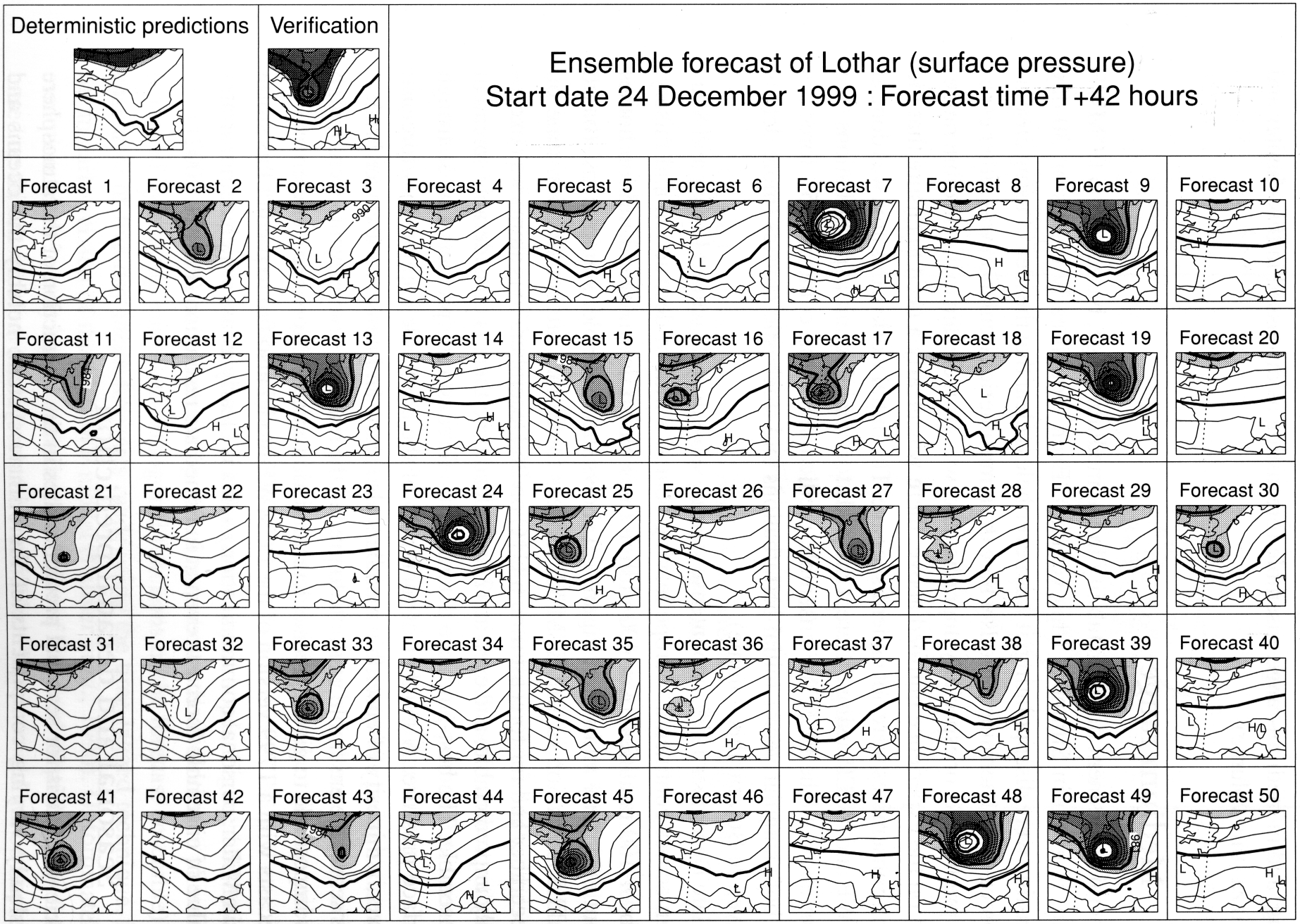
Follow another close-by set of initial conditions within the ball. Now we see more rapid growth of errors.



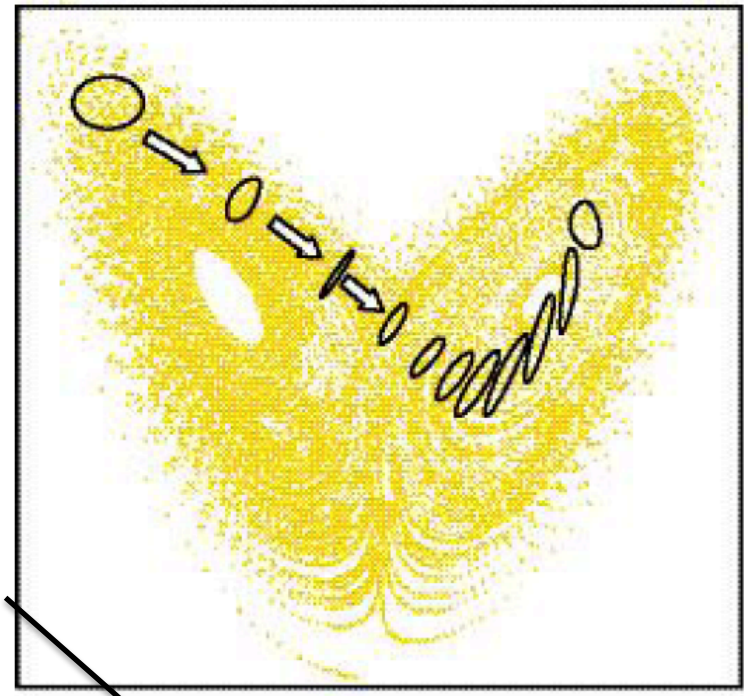


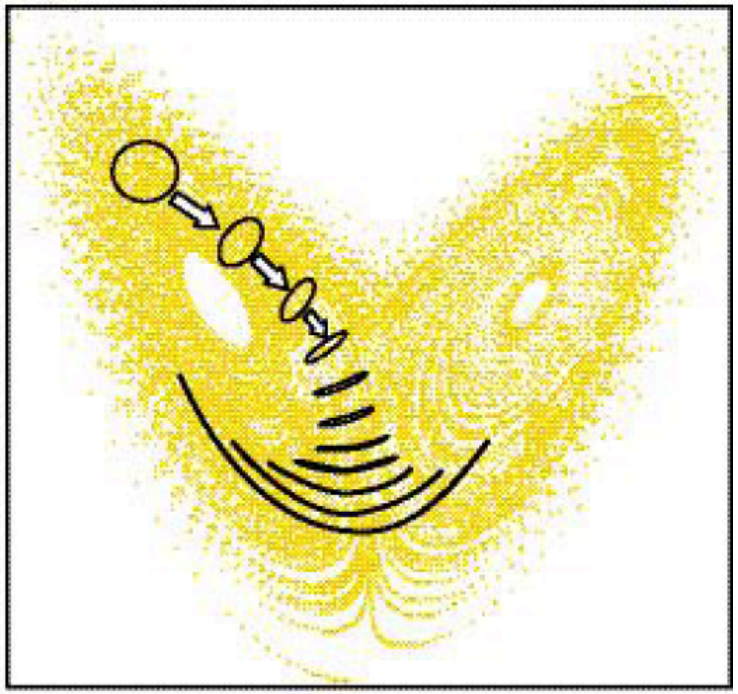
Follow this set of initial conditions within the ball. Here, an extremely rapid growth of errors.

Weather forecasts can exhibit this extreme error growth, too.



c/o Tim Palmer

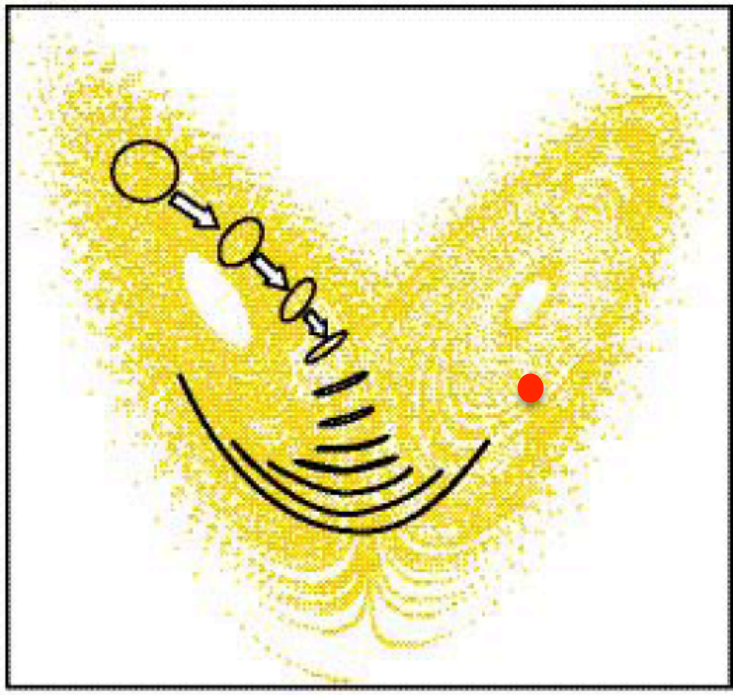




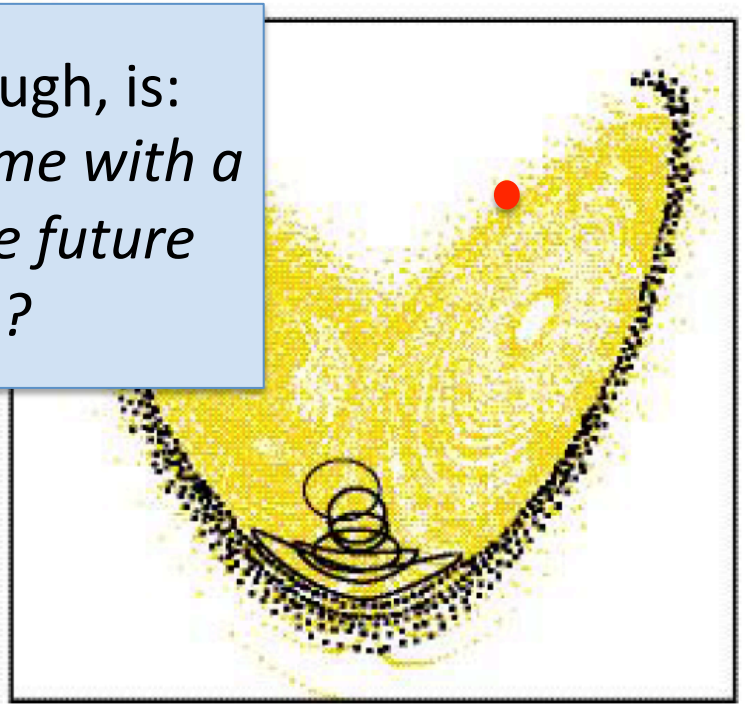
Lessons:

Trying to determine which ensemble solution is most likely isn't a good use of your time.

Prepare your users for (at least) the variety of weather that the ensemble tells you.

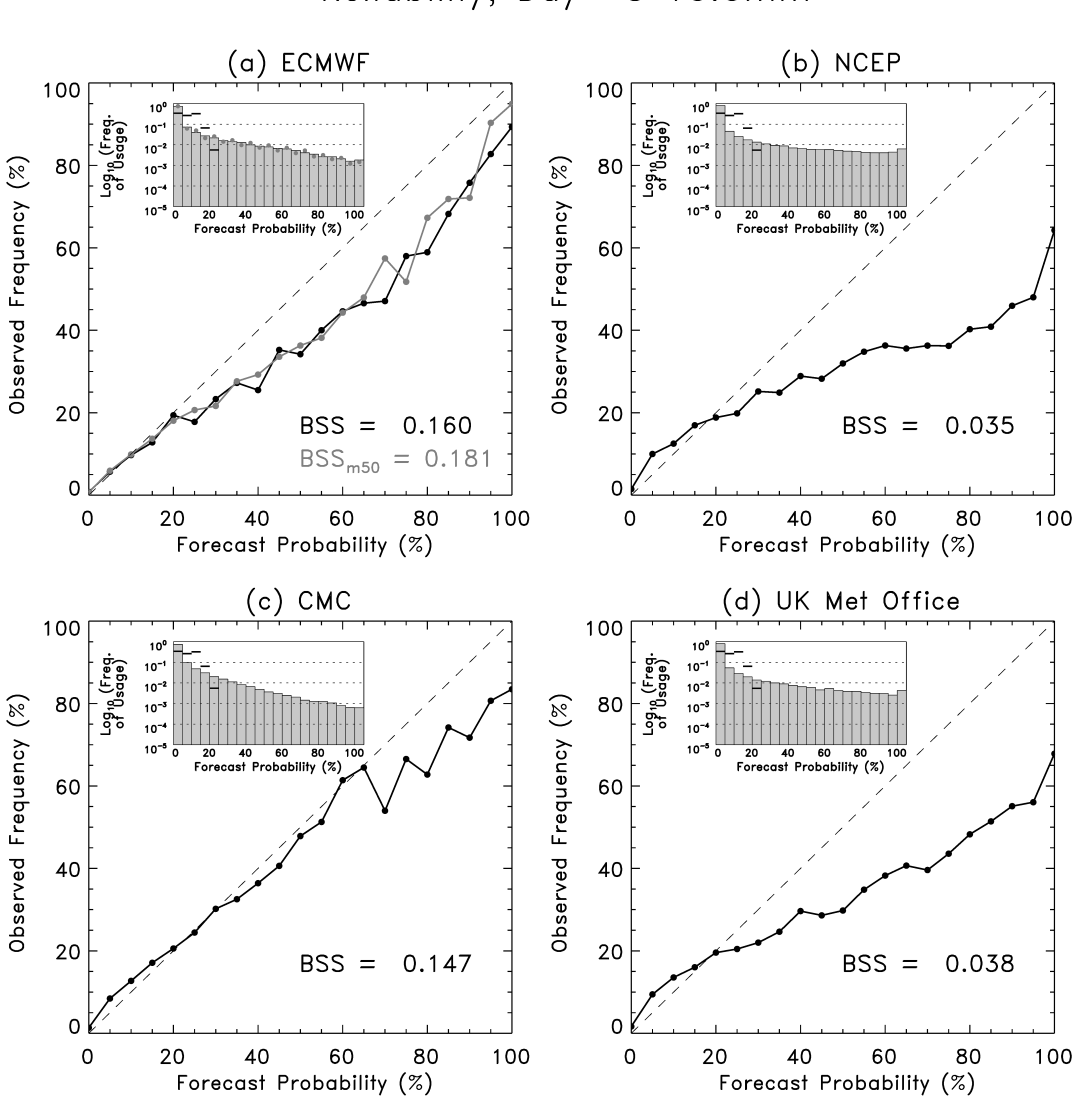


A good thing to ask, though, is:
Is my ensemble providing me with a realistic range of possible future weather scenarios?



Suppose the observed was at the red dot in each case.

Most of our current ensemble systems have this problem, to varying degrees.



Here are precipitation forecasts from four different global ensemble prediction systems. Some are more "reliable" than others. Some are more "sharp" than others (they issue more low and high probabilities).

All are overconfident to some degree – you say 100% probability, but it only happens 80% of the time.

The difference between a lower- and a higher-quality ensemble prediction can matter.

Autumn snowstorm wallops Rockies, Plains

Updated 10/30/2009 3:37 AM | Comments 170 | Recommend 48

E-mail | Print | RSS

Enlarge By Bryan Oller, AP

Lon Rust clears snow from the sidewalk in front of his business in Green Mountain Falls, Colo., on Wednesday.

WINTER STORMS

By Doyle Rice, USA TODAY

A record-setting snowstorm that dumped nearly 4 feet of snow across parts of the Rockies by Thursday will threaten parts of the Midwest and South today with heavy rain and flooding.

The powerful fall storm forced hundreds of flights to be canceled in Denver and closed schools and major highways. Heavy snow fell as far west as northern Utah's Wasatch Front to western Nebraska's northern border. In South Dakota, snow shut down the Mount Rushmore National Memorial. A blizzard warning was in effect until morning in northwest Kansas.

The heaviest snow was reported in the foothills west of Denver, near Pinecliffe, Colo., with 43.8 inches by midday Thursday, according to the National Weather Service. October snow records were set for Cheyenne, Wyo., and Fort Collins, Colo.

Share

Add to Mixx

Facebook

Twitter

More

Subscribe

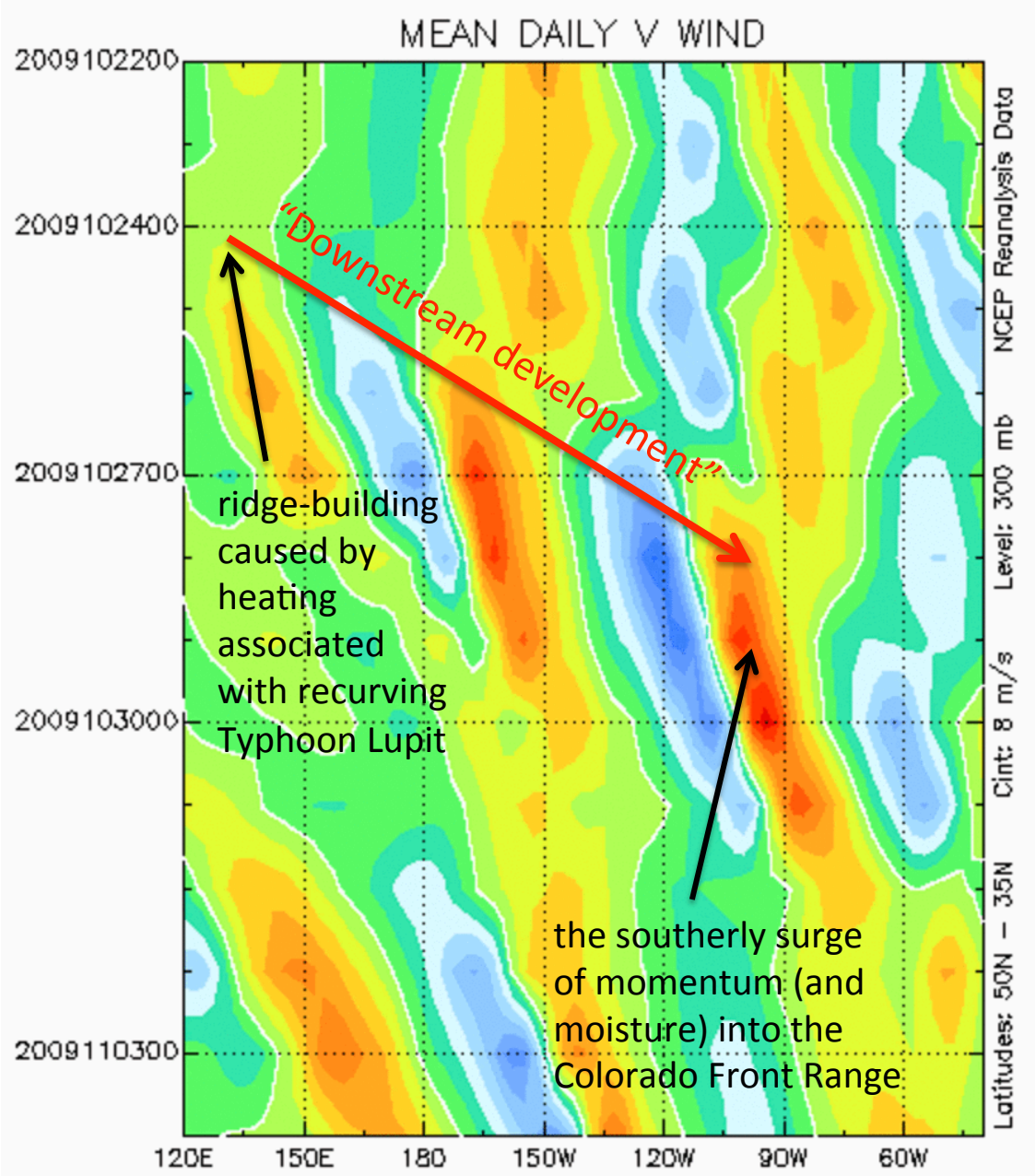
myYahoo

iGoogle

More

USA Today, 30 Oct 2009

Here's the sort of high-impact event it would be extremely useful to be able to have advanced warning of at the weather-climate timescales.

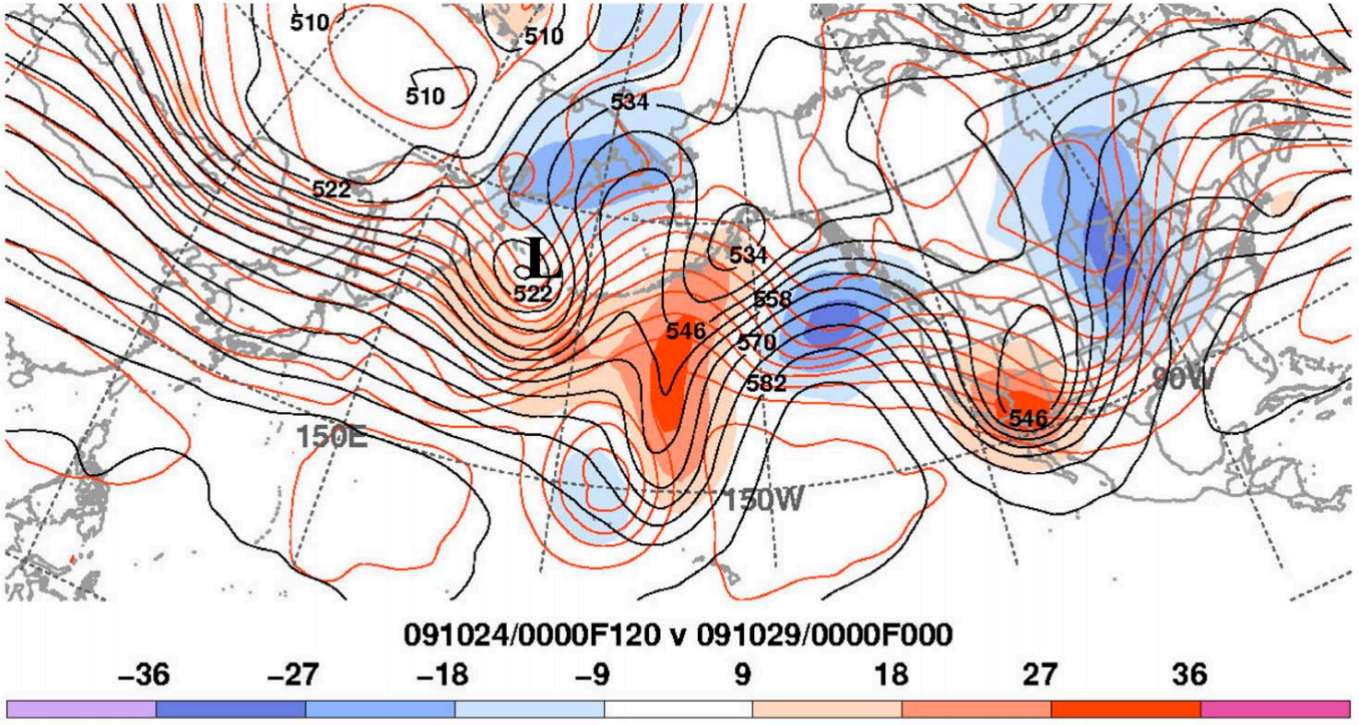


Hovmöller of 250 hPa meridional wind analyses

In 3.5 days, the wave train has covered a third of the globe.

Use of a global forecast model is clearly necessary, if not sufficient.

Massive bust of day +5 NCEP global deterministic; totally misses jet stream pattern



500-hPa geopotential height (black, dam), 120-h forecast (red, dam), and 120-h Forecast − Analysis (shaded, dam)

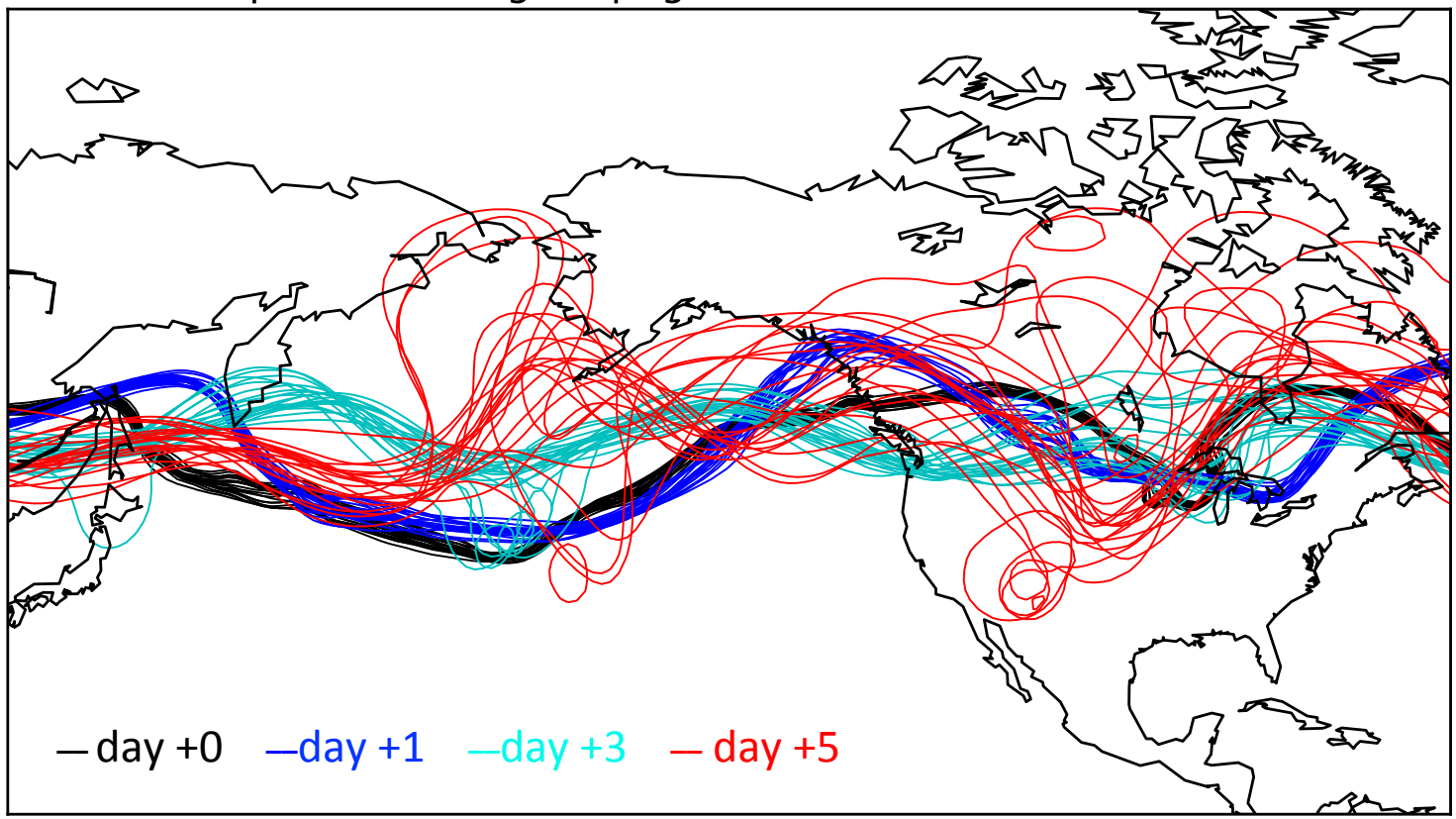
Perhaps small initial condition errors led to the bust?
What about ensemble systems?

c/o Heather Archambault, SUNY/Albany and NRL/Monterey

“Spaghetti plots”, NCEP ensemble

(546 dam contour, + 5 day forecast)

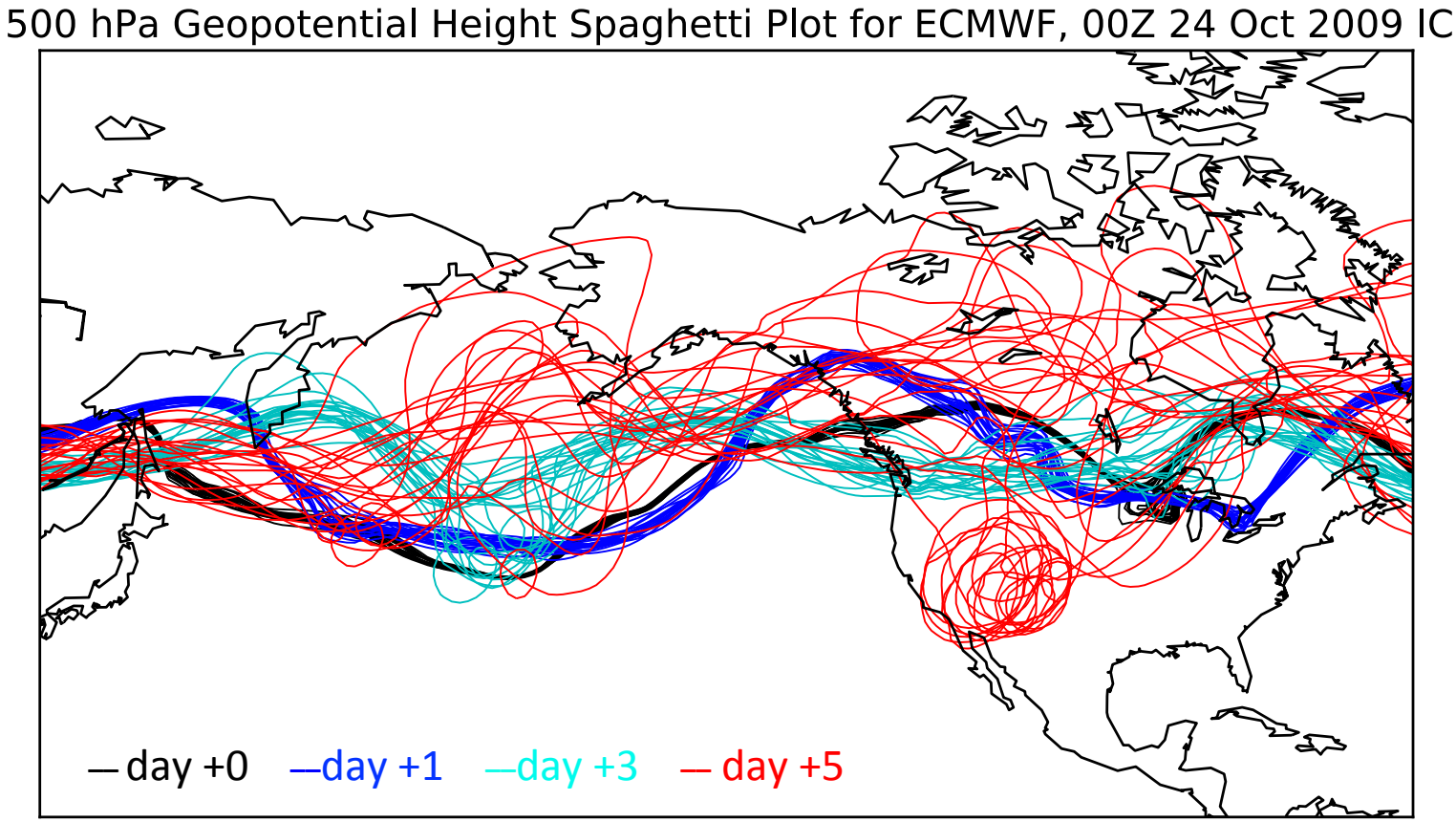
500 hPa Geopotential Height Spaghetti Plot for NCEP, 00Z 24 Oct 2009 IC



NCEP ensemble has only a hint in a few members of a major system affecting the southwest US.

“Spaghetti plots”, ECMWF ensemble

(546 dam contour + 5 day forecast)



ECMWF system much better at predicting event in central Rockies. Lessons:

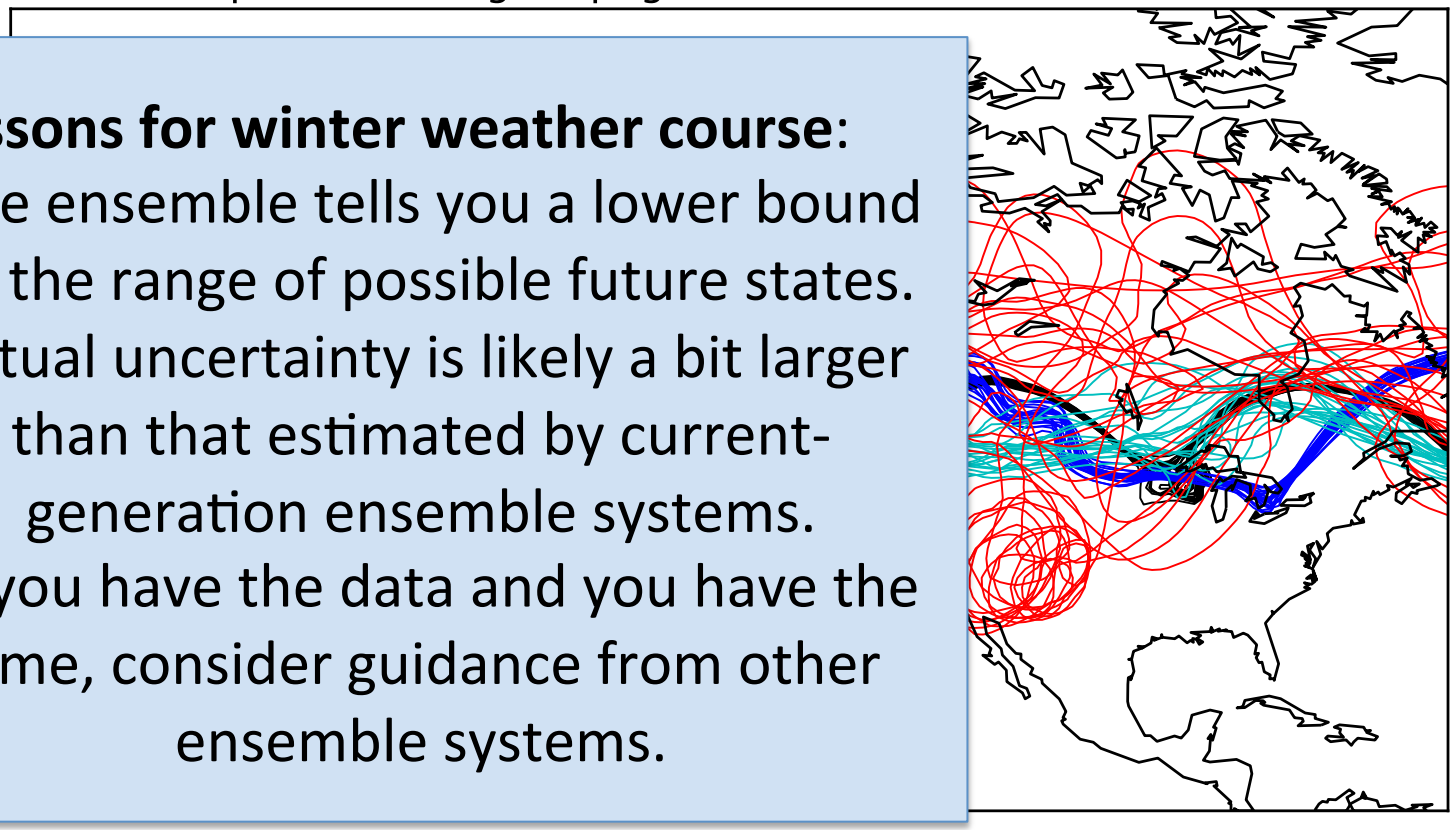
(1) Probabilistic, not deterministic forecasts, are definitely needed.

(2) A high-quality ensemble prediction system is a necessity for weather-climate prediction.

“Spaghetti plots”, ECMWF ensemble

(546 dam contour + 5 day forecast)

500 hPa Geopotential Height Spaghetti Plot for ECMWF, 00Z 24 Oct 2009 IC



Lessons for winter weather course:

(1) The ensemble tells you a lower bound on the range of possible future states. Actual uncertainty is likely a bit larger than that estimated by current-generation ensemble systems.

(2) If you have the data and you have the time, consider guidance from other ensemble systems.

ECMWF system much better at predicting event in central Rockies. So:

(1) Probabilistic, not deterministic forecasts, are definitely needed.

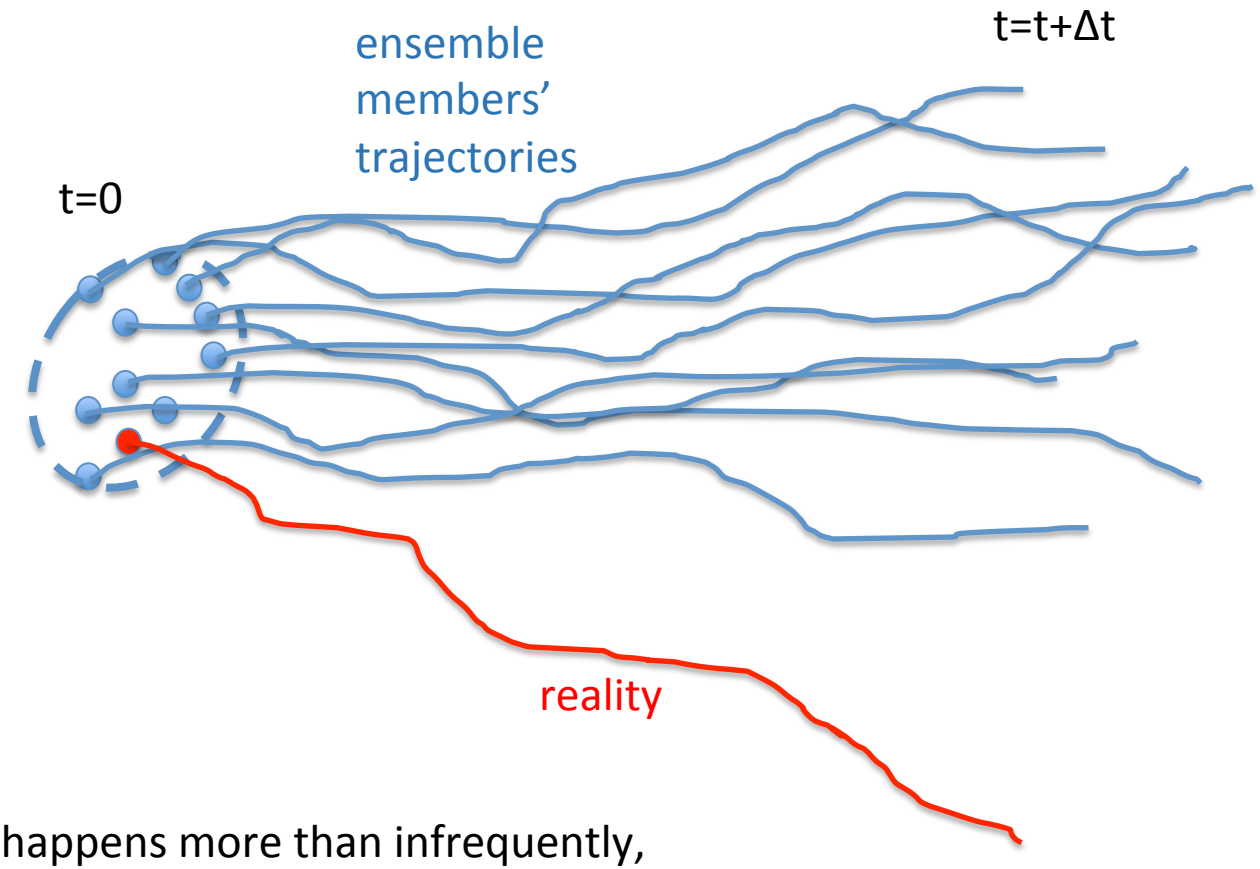
(2) A high-quality ensemble prediction system is a necessity for weather-climate prediction.

Ways to create ensembles (not mutually exclusive)

- Multiple deterministic forecasts: "poor-man's ensemble"
- Vary the initial conditions ("EnKF," "ETR," "Singular vector," "bred modes," "multi-analysis" – we'll discuss principles in a moment) and generate an ensemble with your forecast model.
- Multiple models.
- Multiple physical parameterizations within a model.
- Stochastic physics – build random processes right into the physical equations of the forecast model.
- With a global model, or with a regional model.

What are the guiding principles?
To use them wisely, it helps to understand right (and sometimes wrong) ways they are generated.

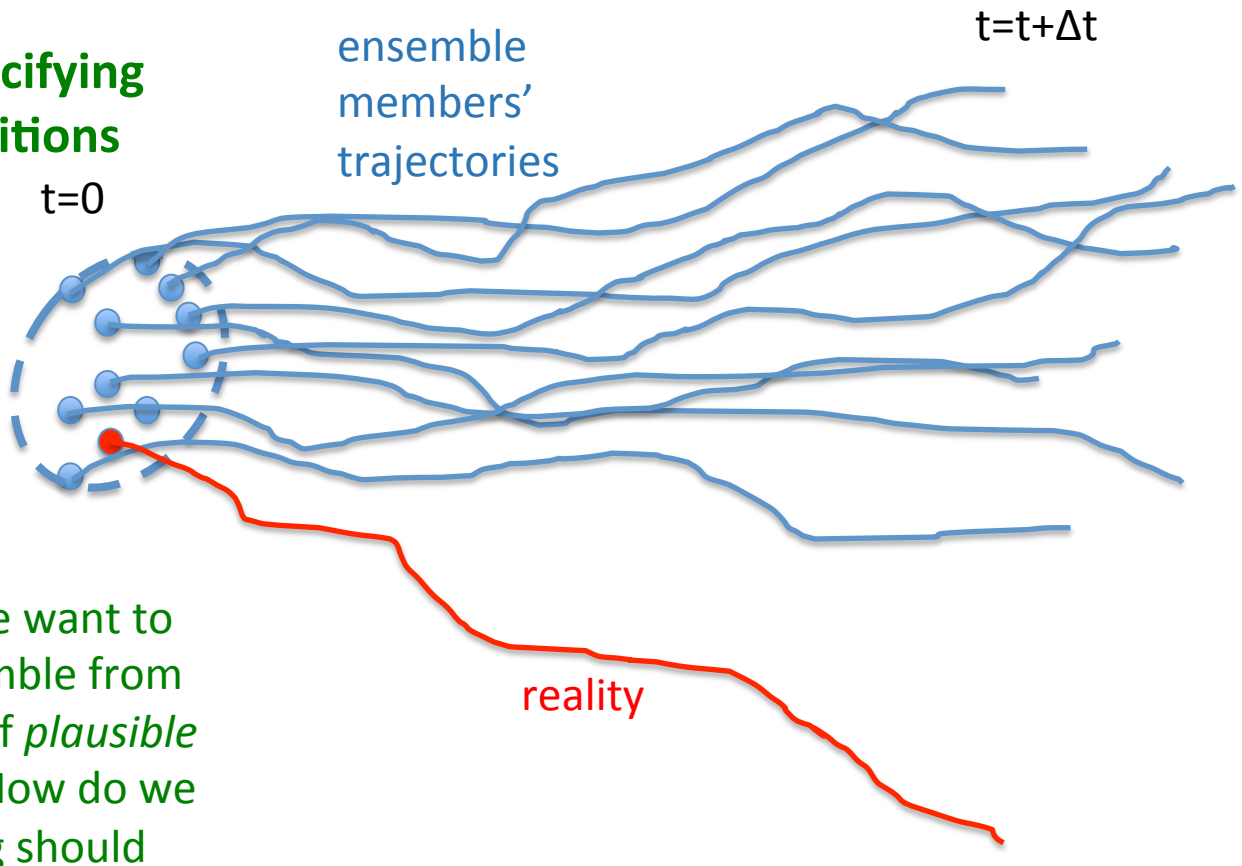
Scientifically, what must be done to produce high-quality ensembles?



If this situation happens more than infrequently,
we need to improve our ensemble prediction system
(that's my job)

Scientifically, what must be done to produce high-quality ensembles?

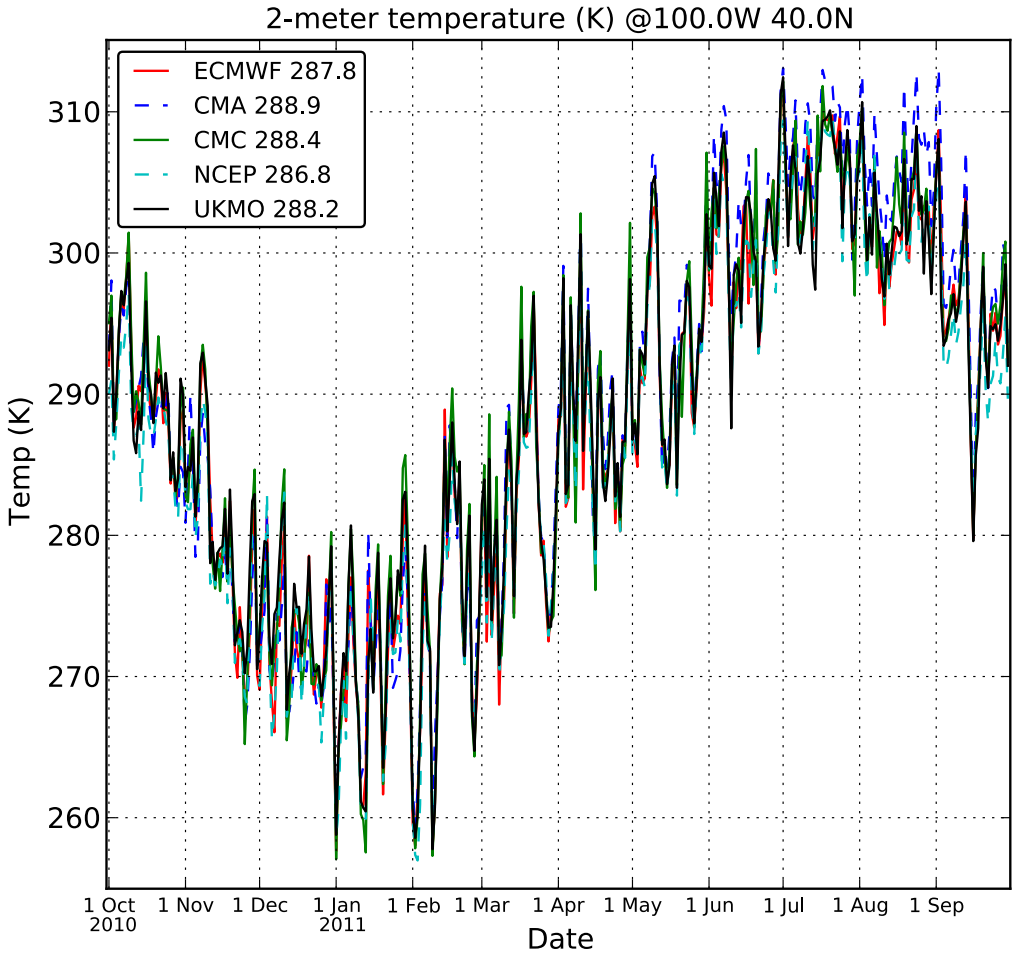
Problem 1: Specifying the initial conditions



Theory tells us we want to sample the ensemble from the distribution of *plausible analysis states*. How do we do that? How big should the differences be. What structure should they have?

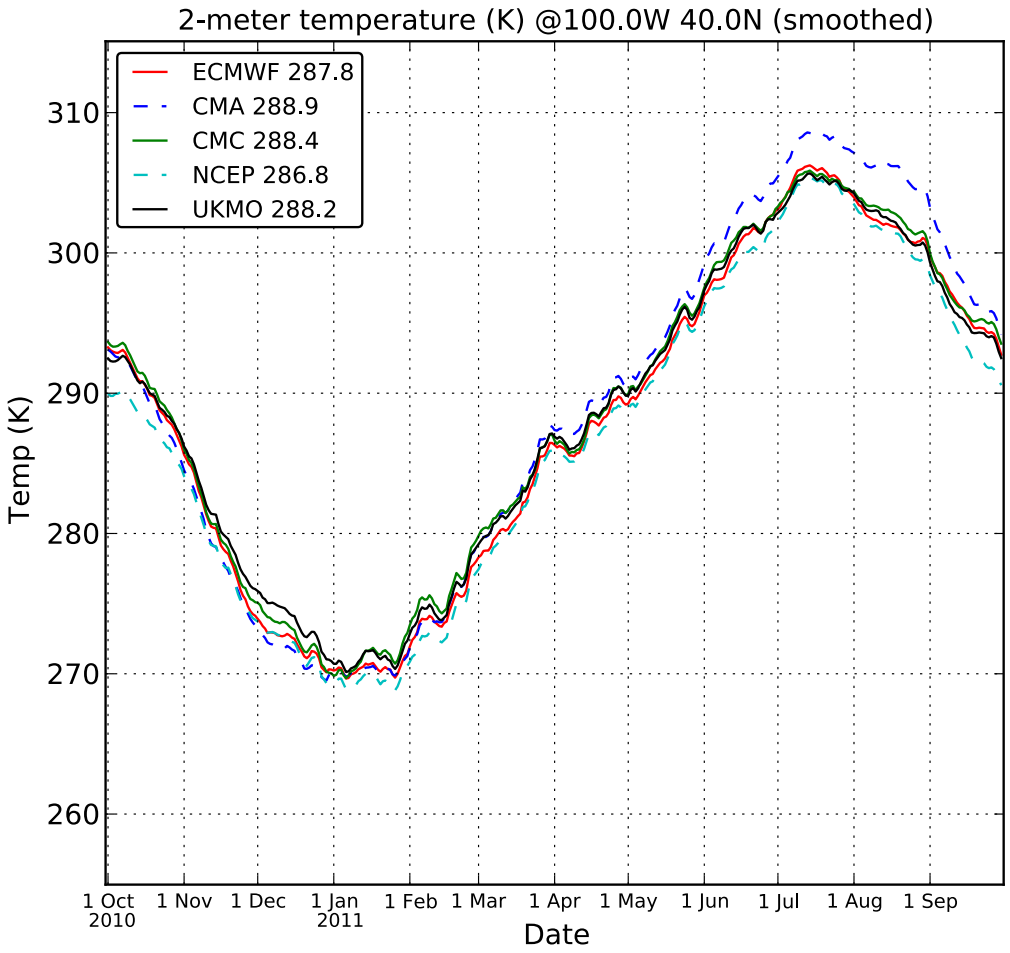
Are analyses uncertain?

A time series of analyses in the central US



Looking at this plot, the various line colors overlap quite a bit, suggesting that the differences have a substantial random component. However, looking at the yearly averaged temperature (listed in the box in the upper left), notice for example that NCEP's analysis is > 2 K colder than CMA's, on average.

Time series of analyses, central US (smoothed)



Here smoothed using average of +/- 15 days.

Warmer CMA analyses in last 4 months stands out.

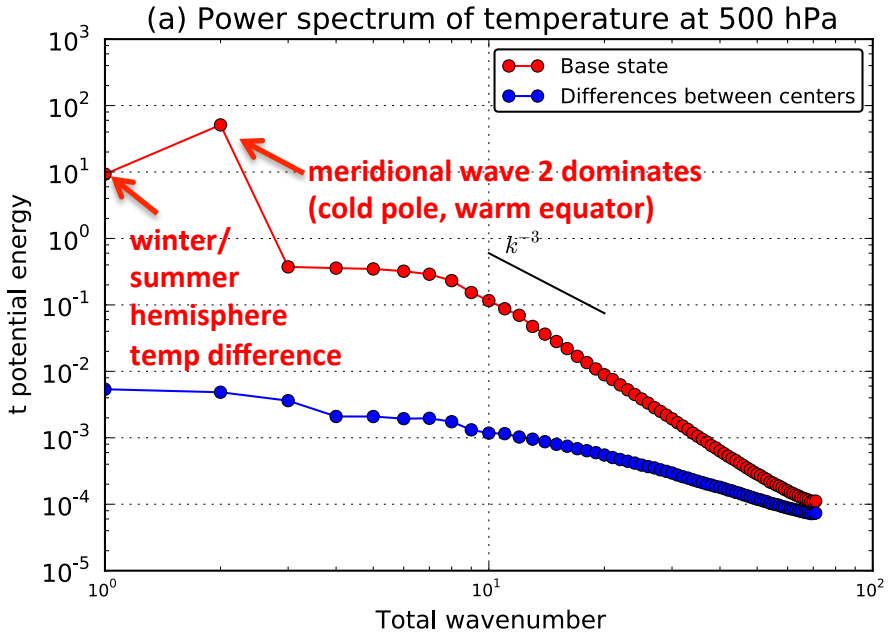
Even in a relatively data-rich region, there are apparent biases in analyses – there isn't agreement on what even the average analyzed temperature should be over the data-rich central US.

Perturbing initial conditions of surface temperature by a degree or two seems warranted.

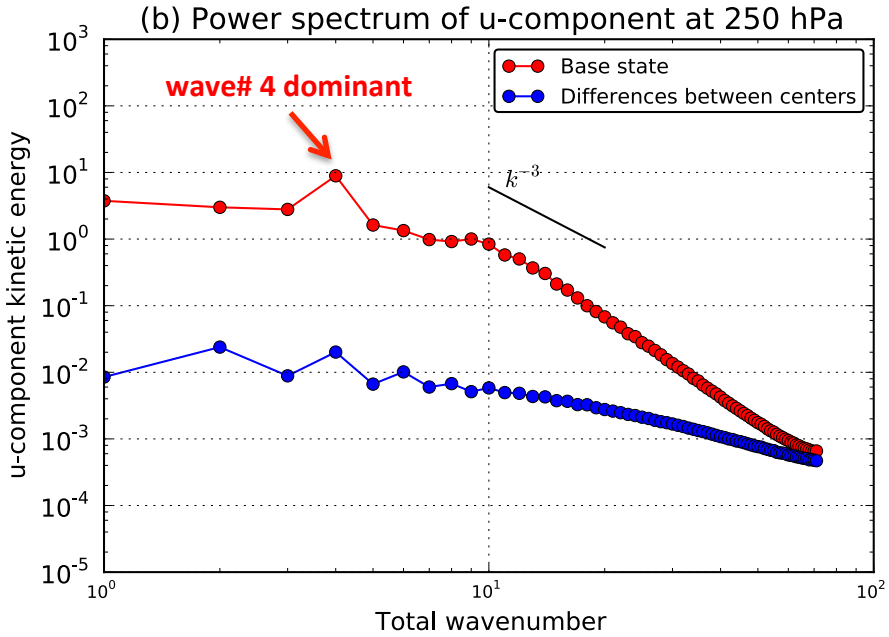
Power spectra from analysis data

ECMWF used for base state; ECMWF - NCEP used for differences, a surrogate for analysis errors. Spectra computed daily, then averaged over the full year.

T @ 500 hPa

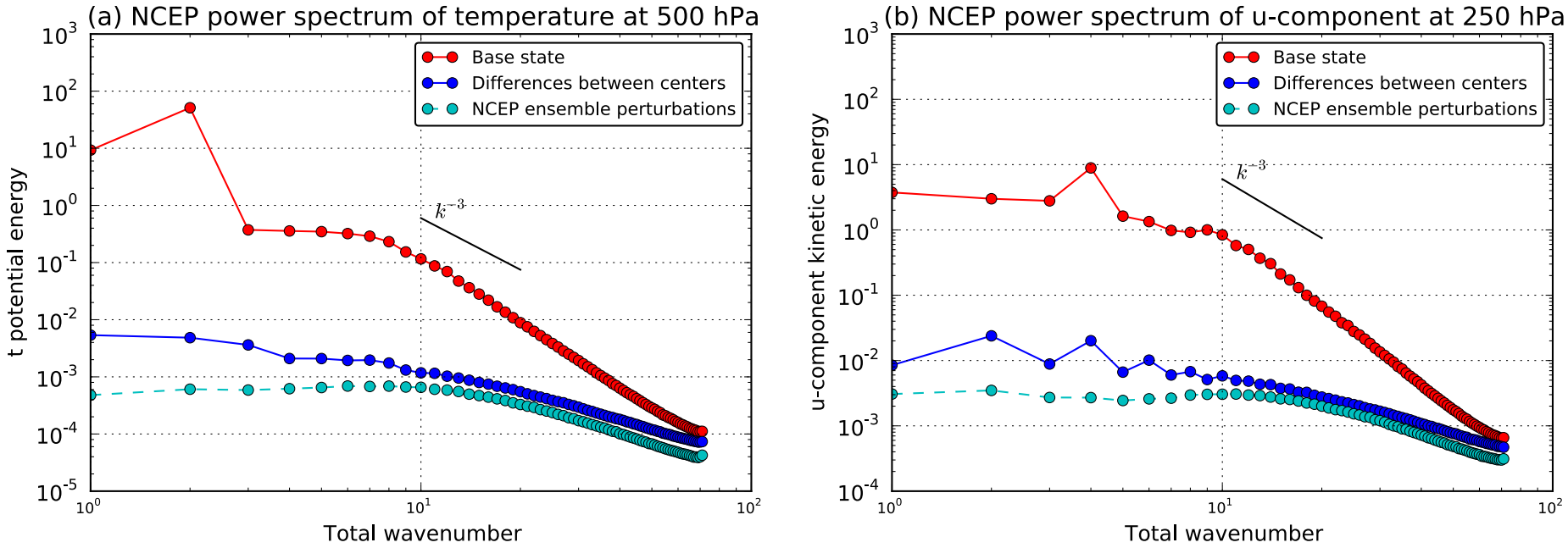


U @ 250 hPa



(1) Larger analysis "errors" (i.e., differences) at larger scales than at smaller scales, but …
(2) Large signal-to-noise ratio (S/N) at large scales, small S/N at smaller scales.
(3) Are analysis errors really that large at the largest scales? Probably yes for some models with larger biases, no for others with smaller biases (e.g., ECMWF).
(4) Analysis errors will have some large-scale correlation structure to them. Not random at every grid point.

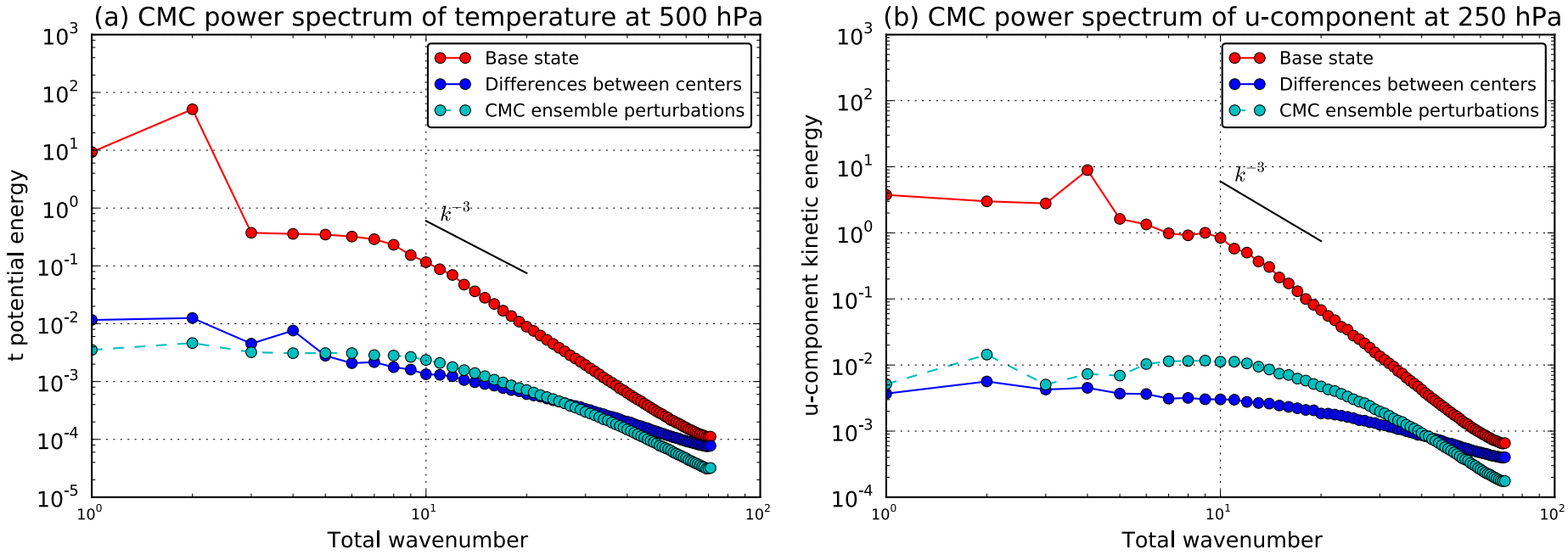
Power spectra of ensemble perturbations, NCEP ensemble



(1) Suggests this ETR perturbation method used at NCEP may have insufficient power at planetary scales. This is consistent with the assumption made in the ETR that the analysis is unbiased while analyses between different centers suggest there is bias.

(2) ETR's underestimate of initial amplitude is the least for the small baroclinic scales. This may be because the breeding method inside the ETR generates perturbations that project onto the (finite amplitude) Lyapunov vectors, dominated by baroclinic scales.

Power spectra of ensemble perturbations, CMC ensemble



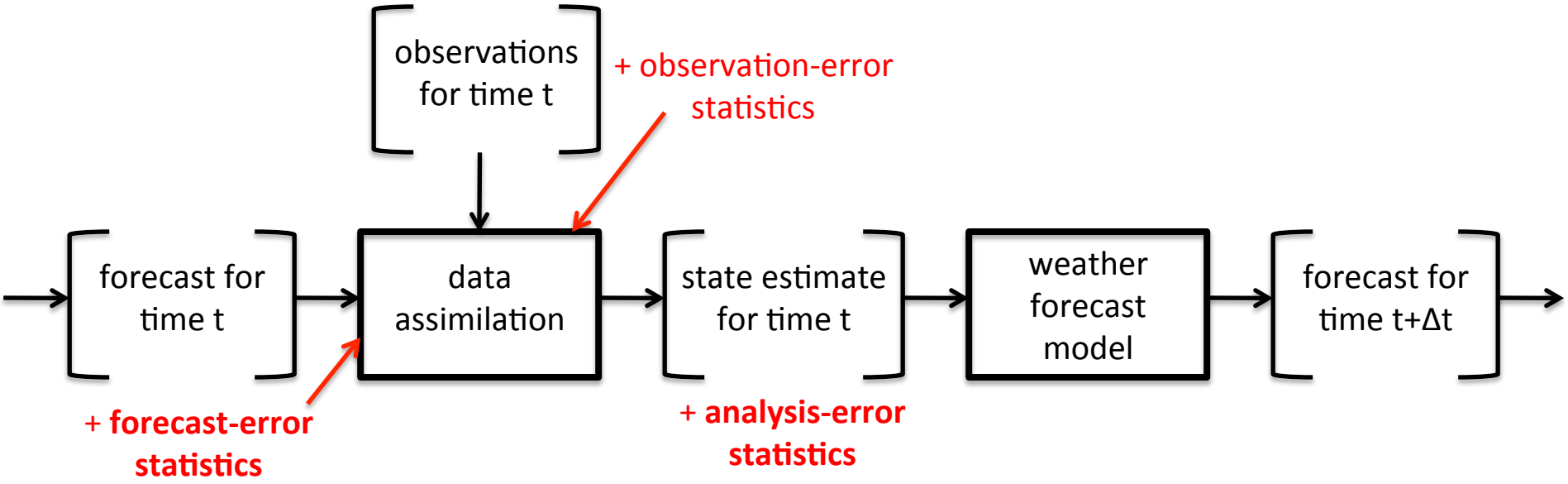
Reasonably well calibrated overall for T500; a bit of an overestimate of variance for U250.

So, we want to generate a range of plausible analyses to initialize the ensemble. What are the principles?

- Initialize from multi-analyses?
  - only a handful of analyses available. What if we want to generate 20 or 50 ensemble forecasts?
  - also: data sharing might be problematic.

Data assimilation:

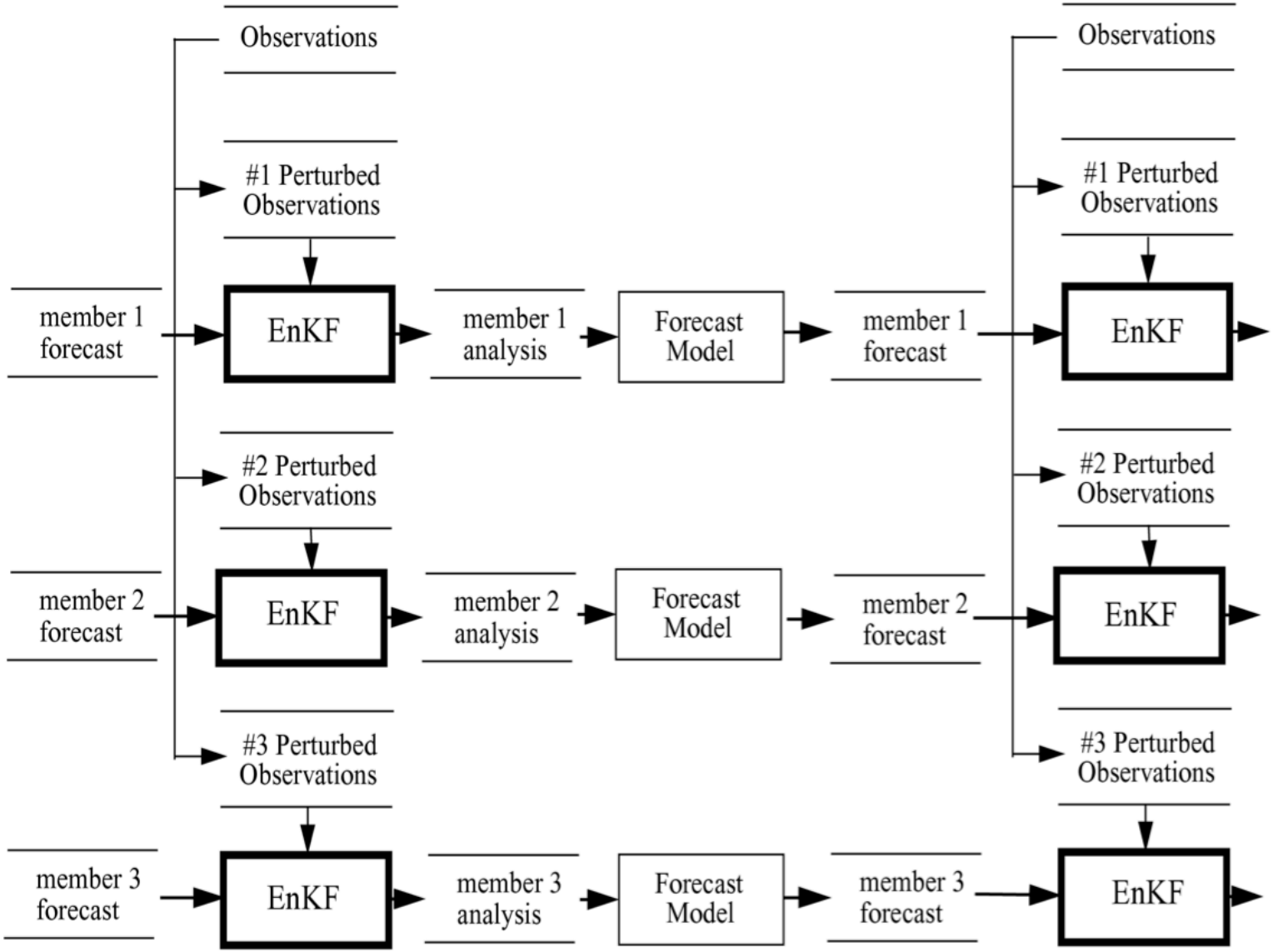
producing multiple possible analyses for initializing an ensemble



To get a reasonable estimate of the state and its uncertainty, we need observations, forecast(s) and we need to simulate the effects of observation-error statistics and **forecast-error statistics**.

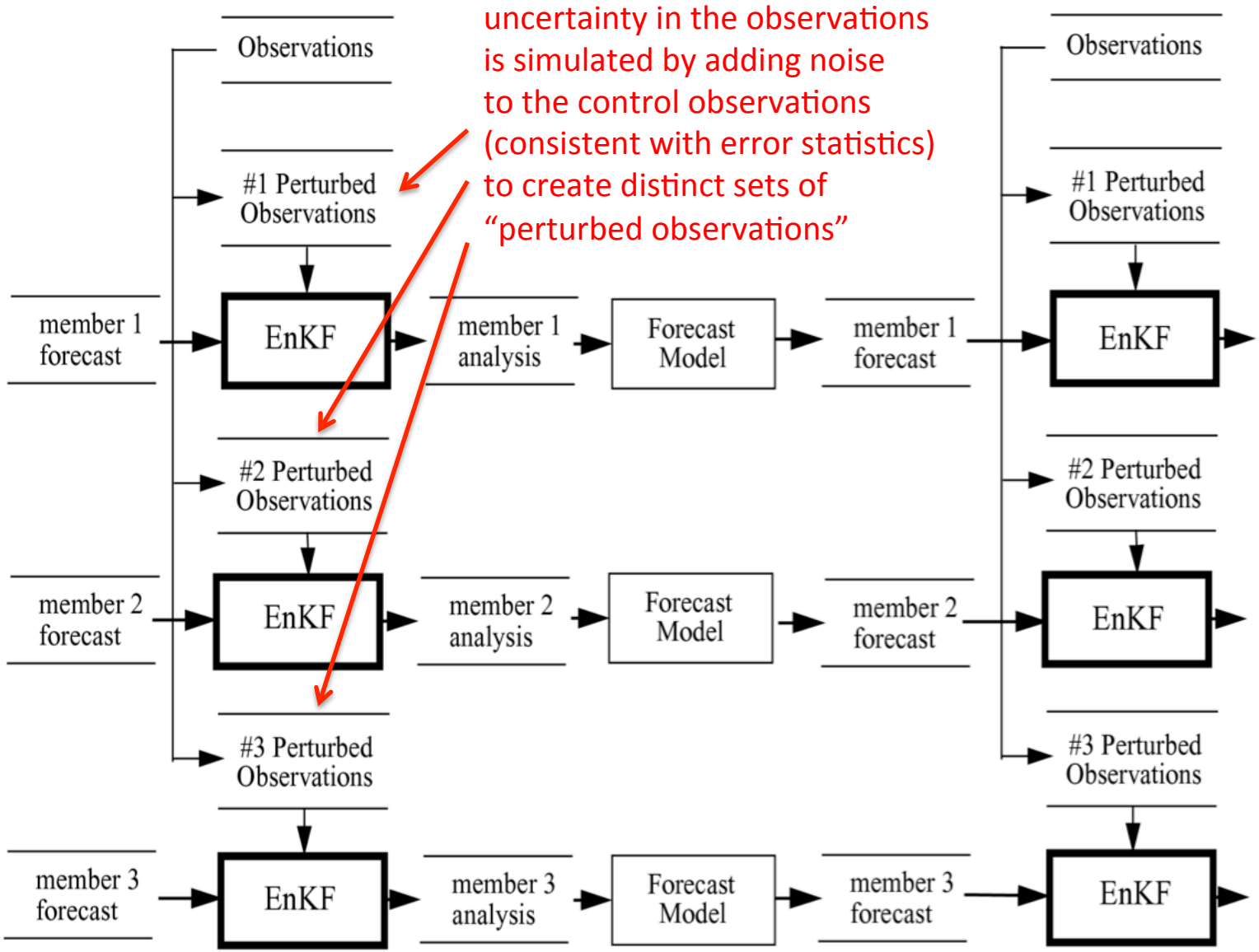
The ensemble Kalman filter: a schematic

(a way to simulate sources of uncertainty in analyses)



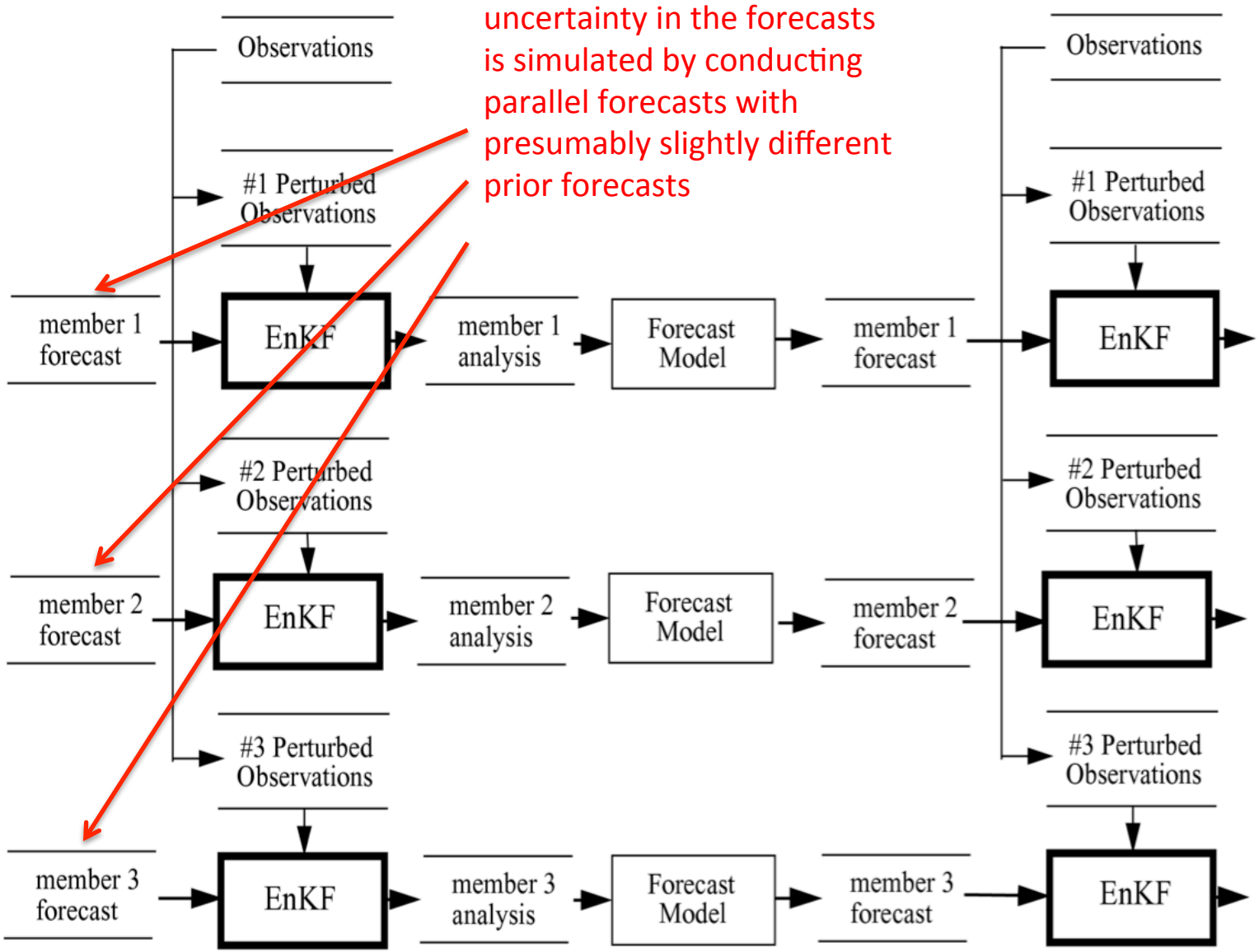
(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

The ensemble Kalman filter (EnKF) : a schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

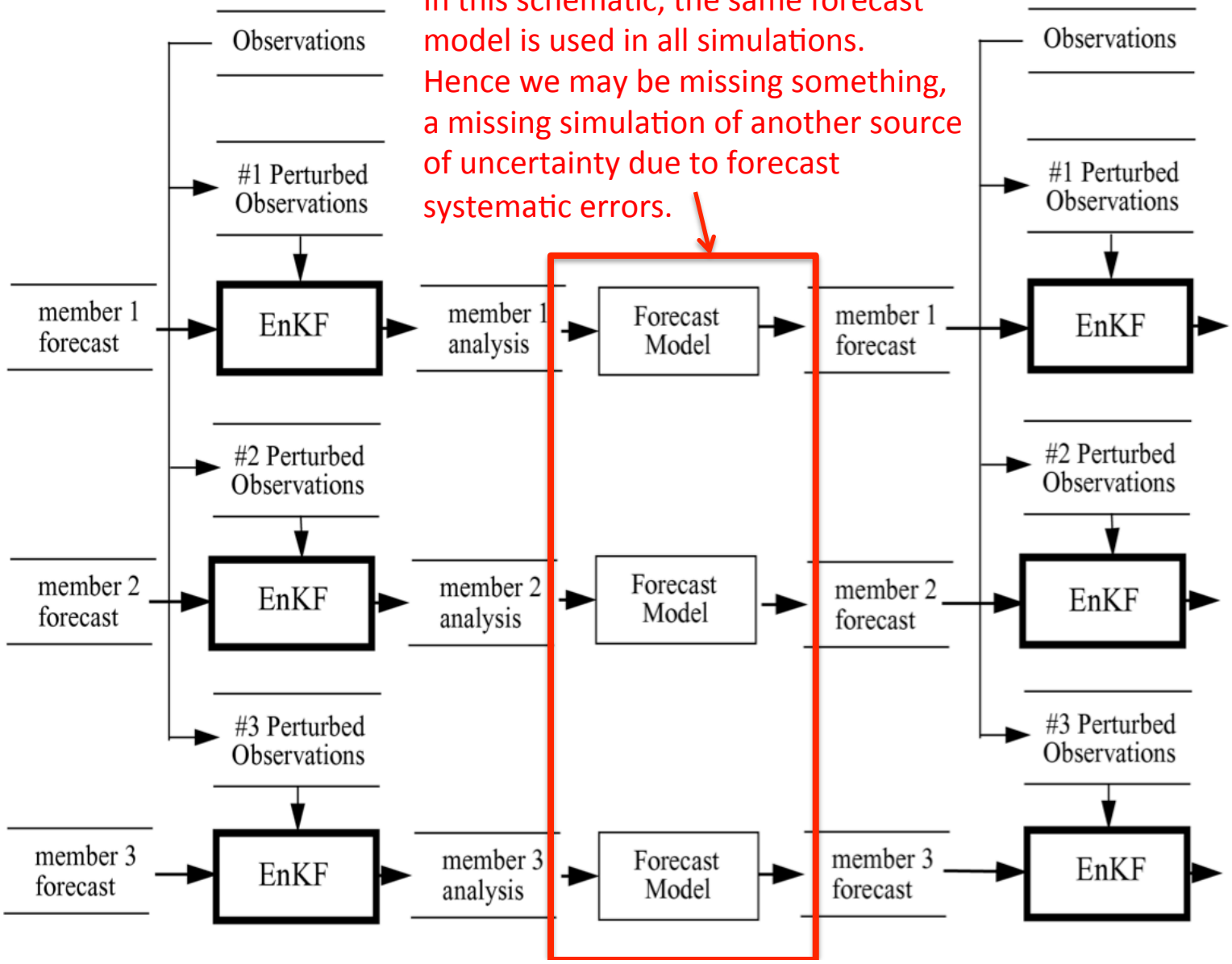
The ensemble Kalman filter (EnKF) : a schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

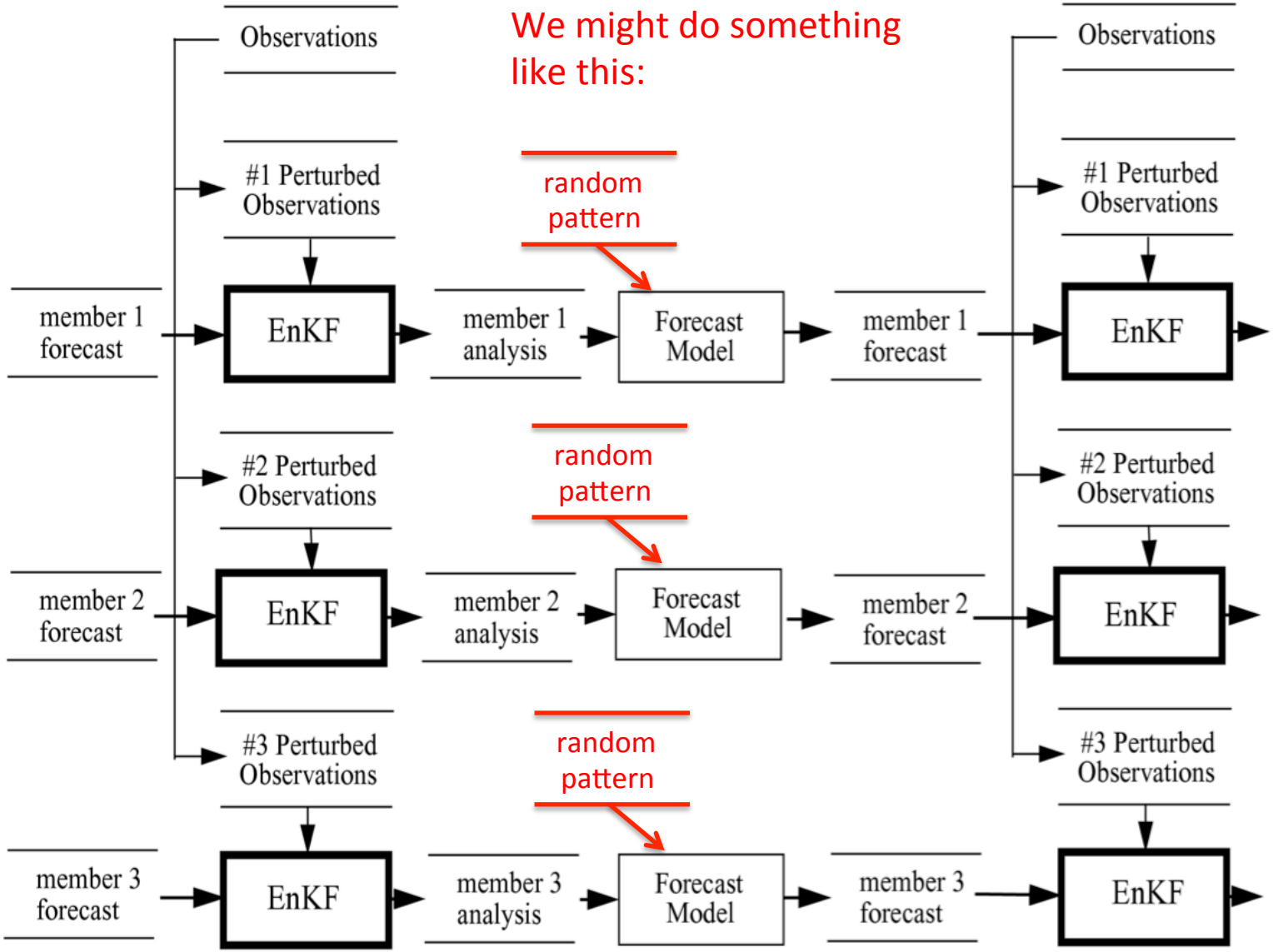
The ensemble Kalman filter (EnKF) : a schematic

In this schematic, the same forecast model is used in all simulations. Hence we may be missing something, a missing simulation of another source of uncertainty due to forecast systematic errors.



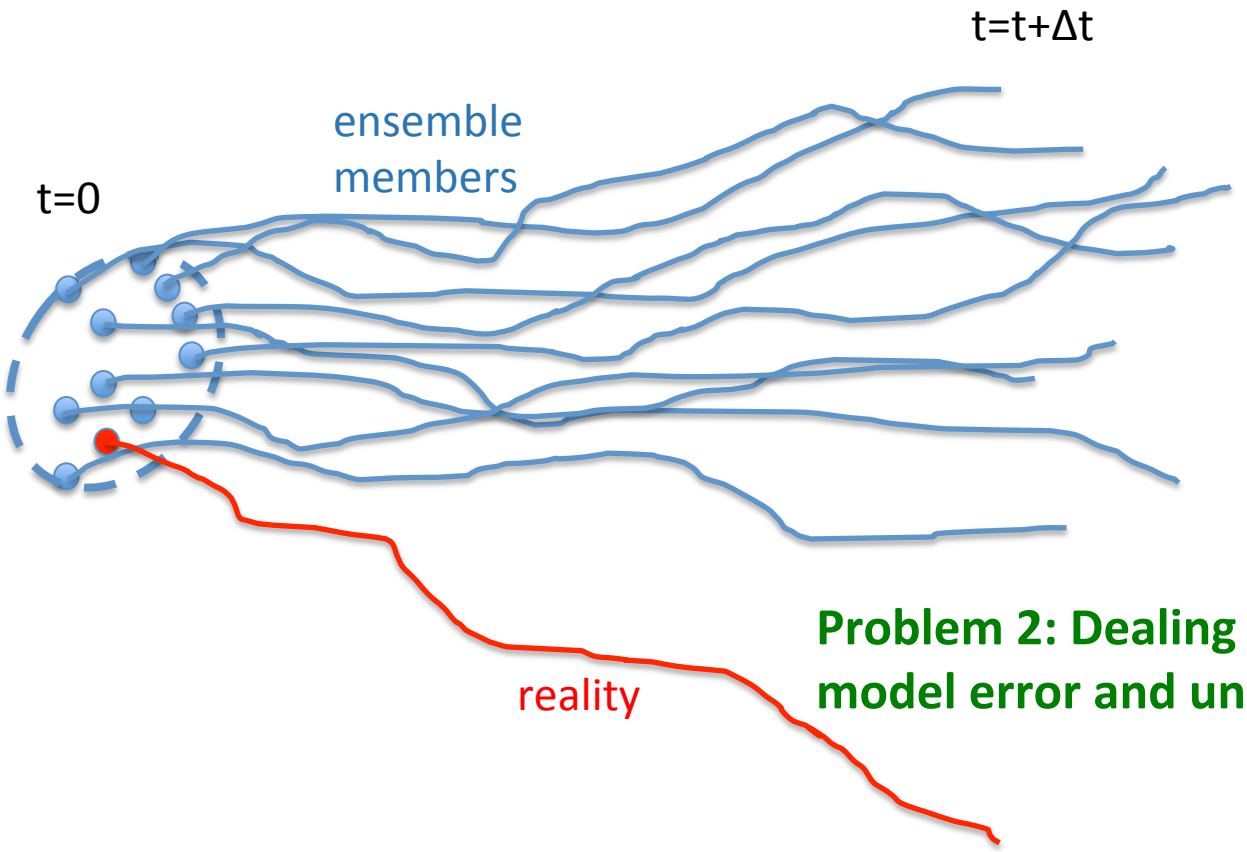
(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

The ensemble Kalman filter (EnKF) : a schematic



(This schematic is a bit of an inappropriate simplification, for EnKF uses every member to estimate background-error covariances)

Scientifically, what must be done to produce high-quality ensembles?



Problem 2: Dealing with model error and uncertainty

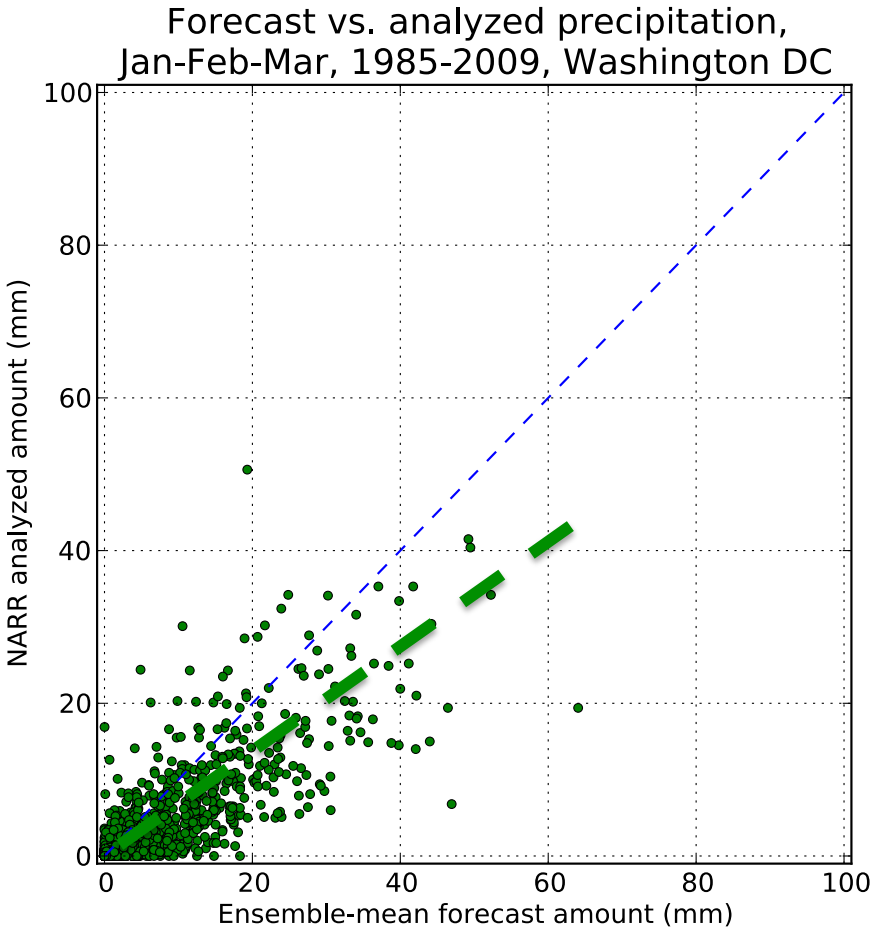
Methods for dealing with model uncertainty

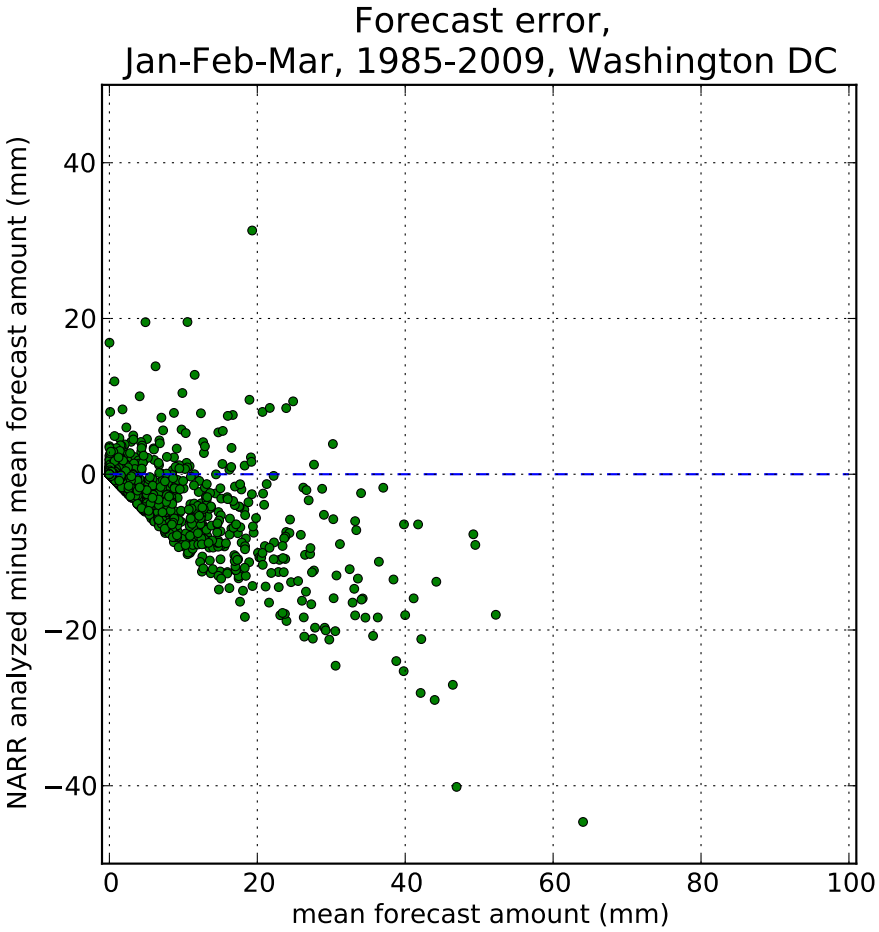
- Make the forecast modeling system better
  - higher resolution, explicit rather than parameterized convection, better observations, better assimilation methods.
- Don't assume your model is perfect:
  - Use multiple forecast models or multiple parameterizations.
  - Add stochastic terms to the forecast model.
- Post-process the guidance. Compare past forecasts to observations/analyses, use this to correct probabilities in the current forecast.

Multi-model vs. statistical post-processing.

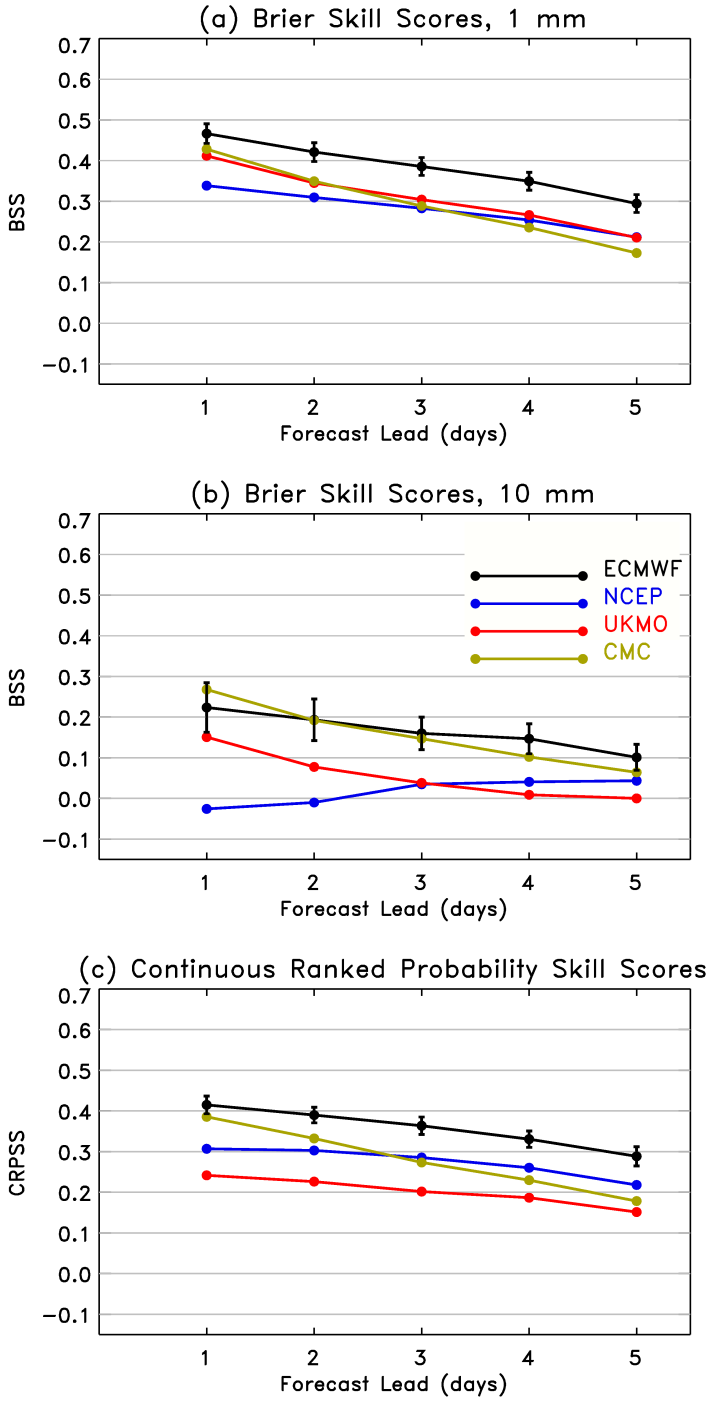
- Let's look at the characteristics of probabilistic precipitation forecasts over the US during 2010.
- Multi-model: 20 members each from NCEP, CMC, UK Met, ECMWF.
- Compare against ECMWF forecasts that have been post-processed using "logistic regression" and 2002-2009 rainfall analyses and ECMWF reforecasts.

Post-processing / reforecasting concept





Run the forecast model for many dates in the past. Use relationship between past forecasts and observations to correct today's forecast. In subsequent slides, a "logistic regression" was used to post-process the probabilities.

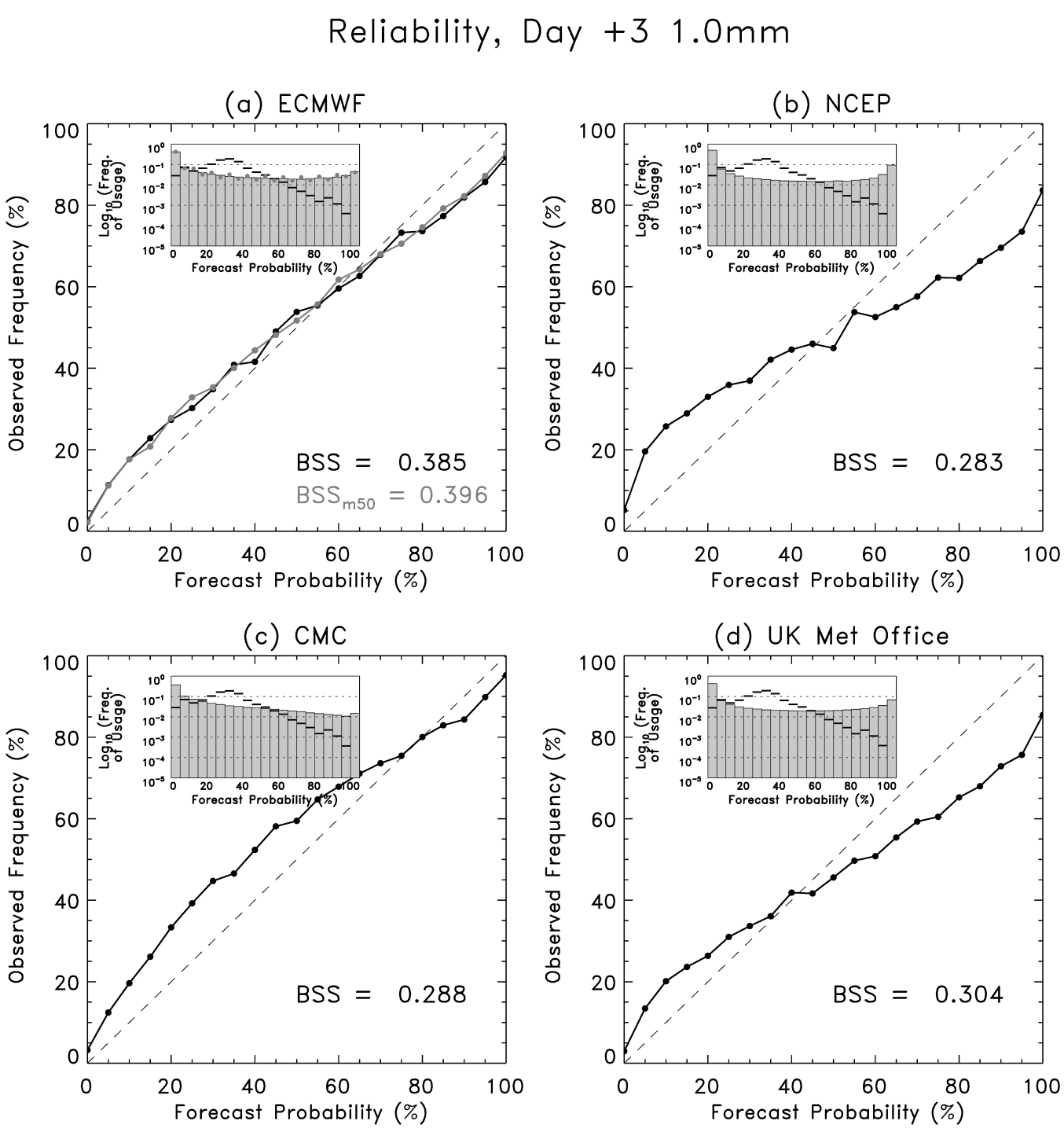


Skill scores of various 20-member ensembles

5th and 95th percentiles using block bootstrap algorithm following Hamill, W&F, 1999.

ECMWF generally the most skillful, though CMC makes similarly skillful 10-mm forecasts.

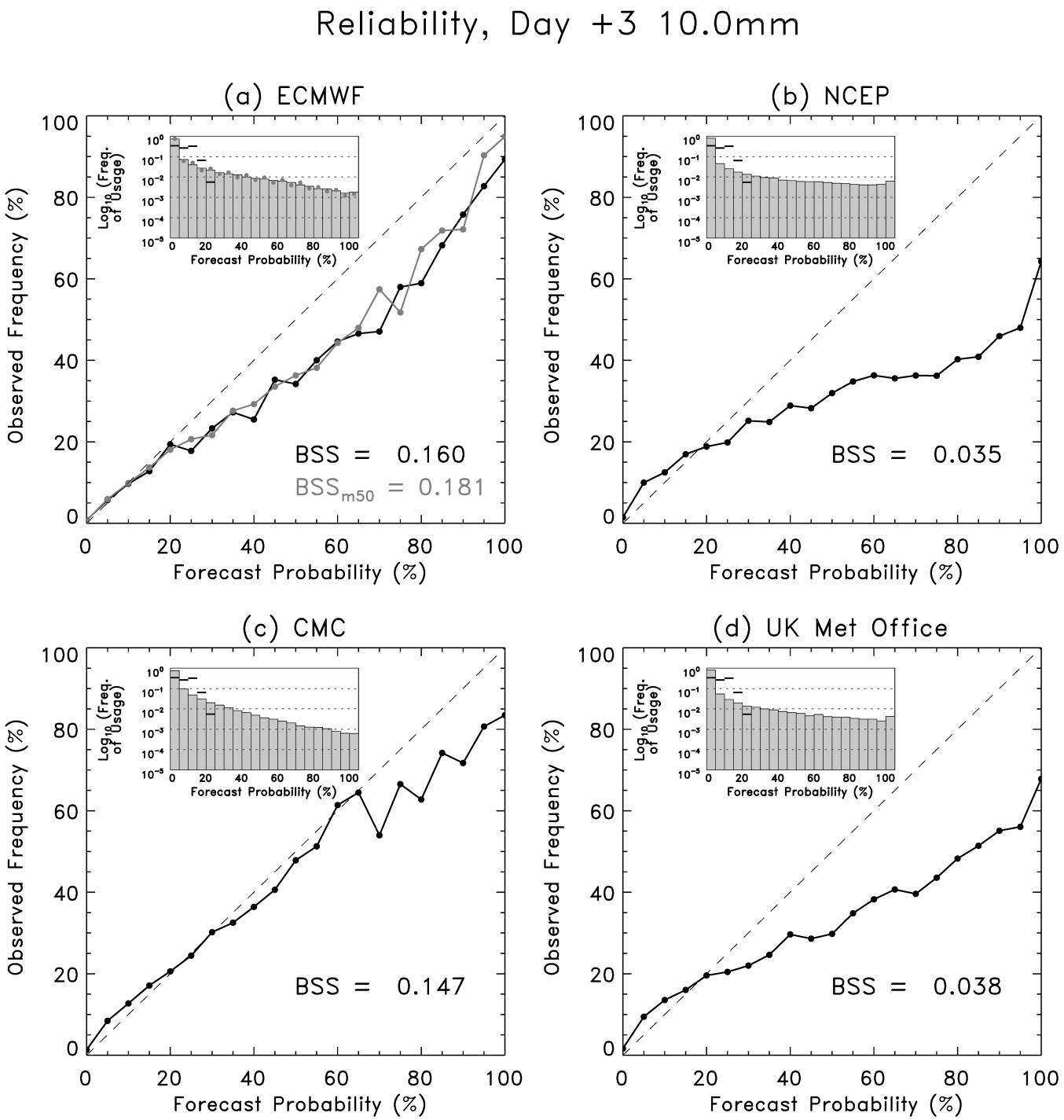
NCEP and UK Met Office trail.



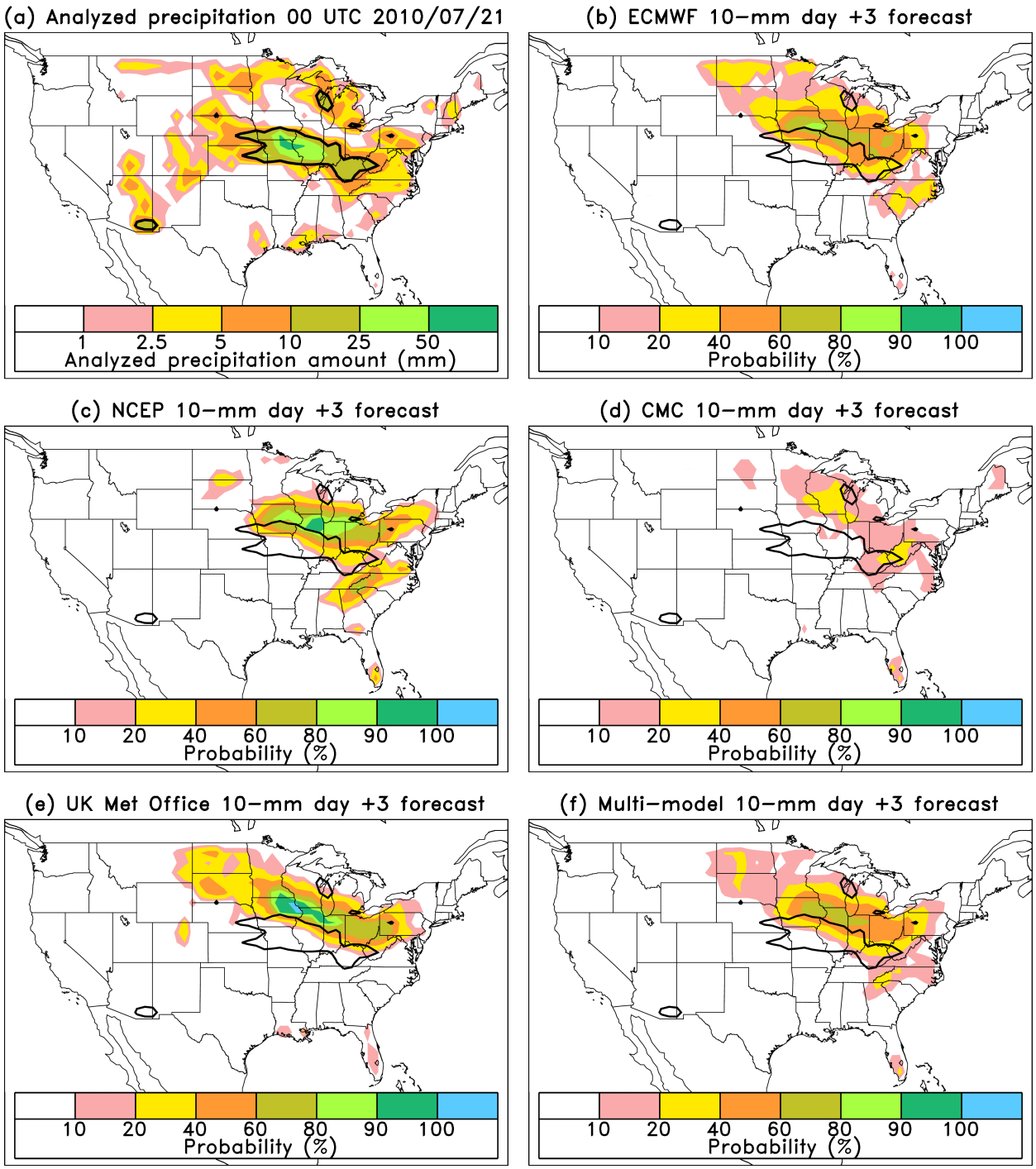
Reliability diagrams, day +3

> 1.0 mm

Inset histogram tells you how often each probability was issued. Black bars for distribution of climatological probabilities for grid points within the CONUS



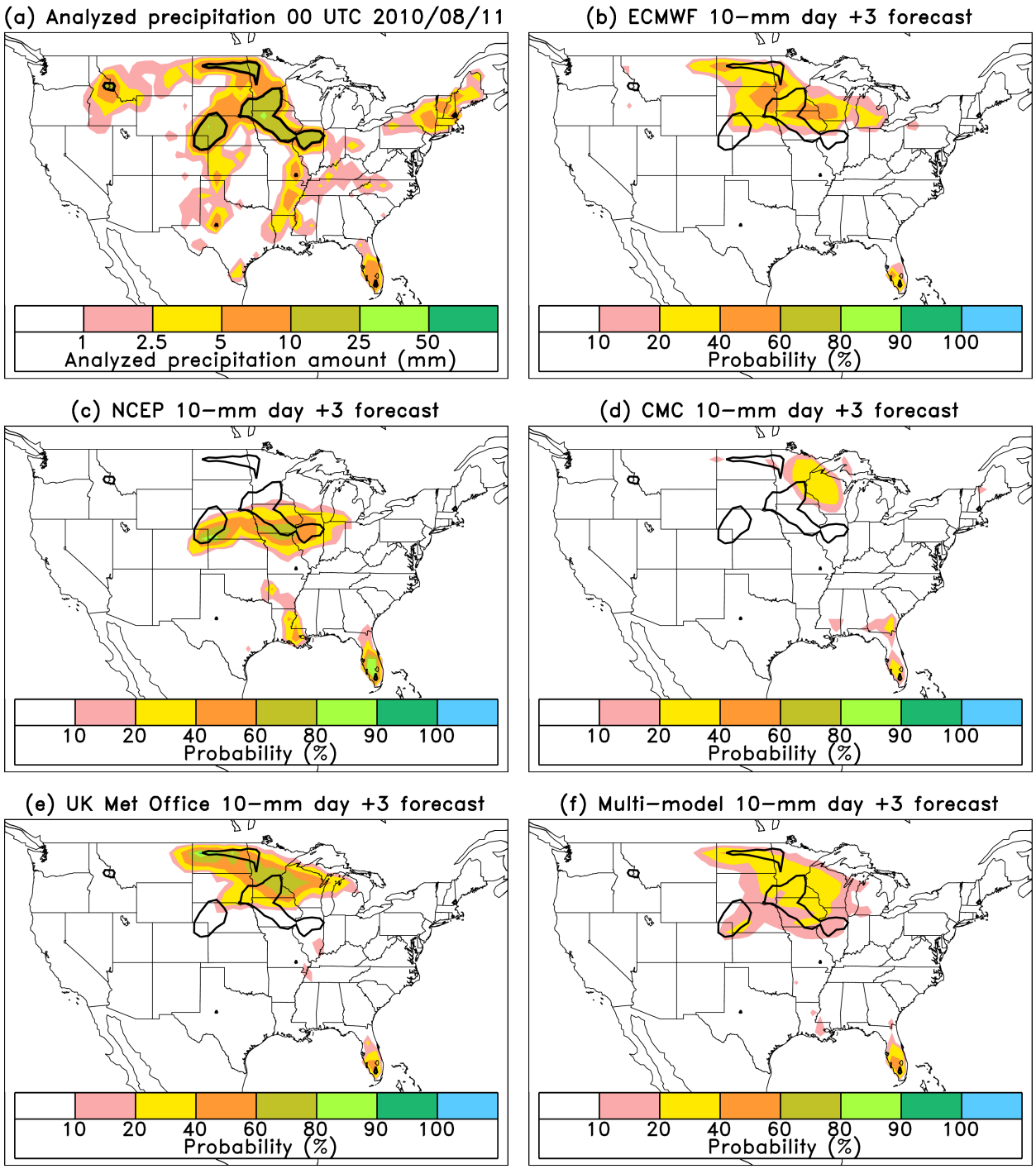
Reliability diagrams, day +3, > 10 mm



Example: where multi-model won't help.

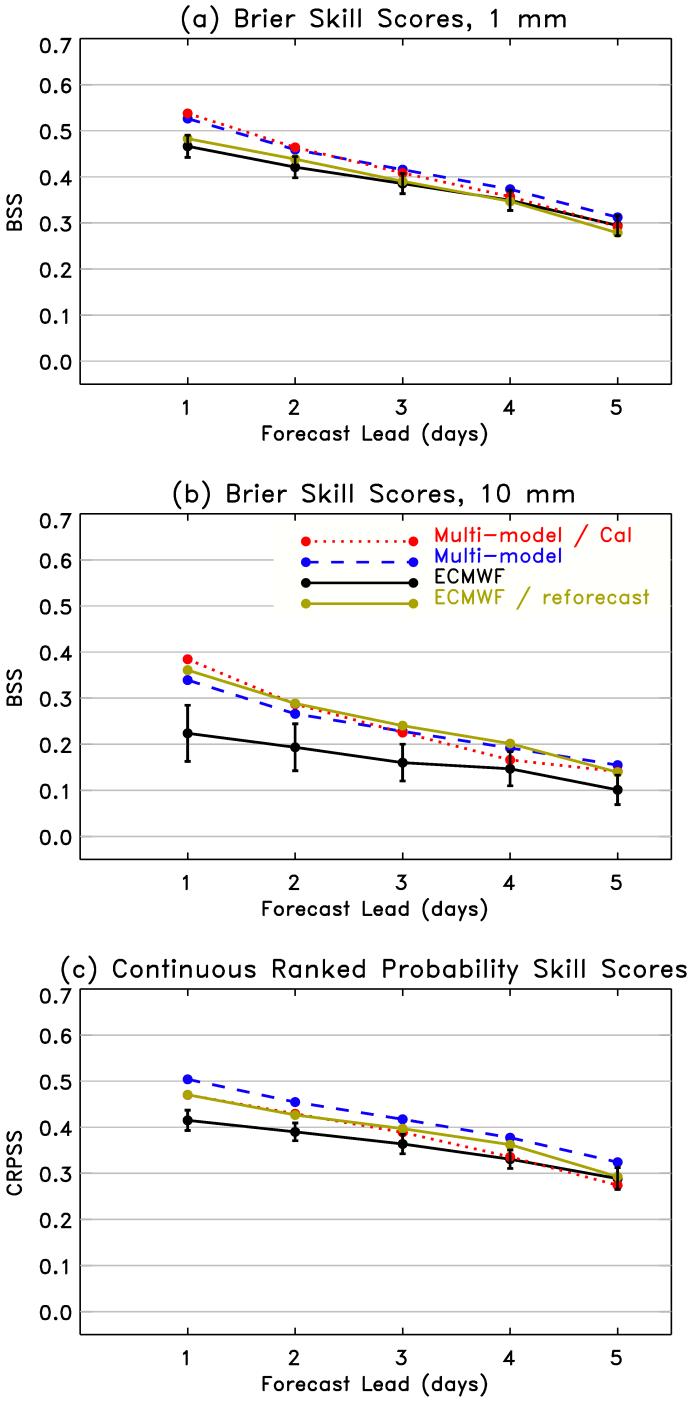
Positional biases are similar in all the models; each is too far north.

(of course, you don't know in advance that the forecast consistency was unrealistic!)



Example: where multi-model should help.

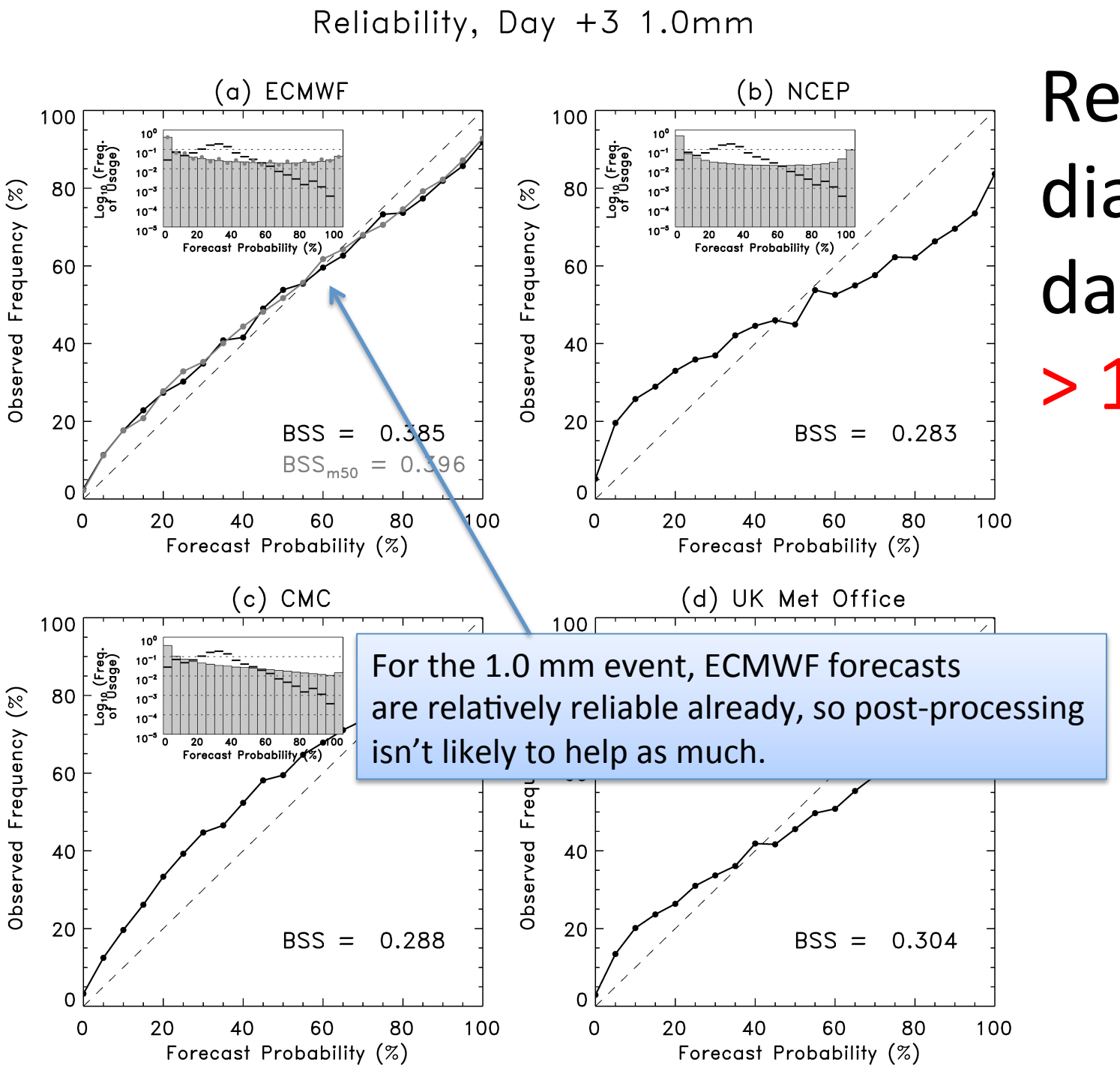
Positional biases are different; NCEP south, ECMWF north.



Skill scores for multi-model and reforecast-calibrated

Notes:

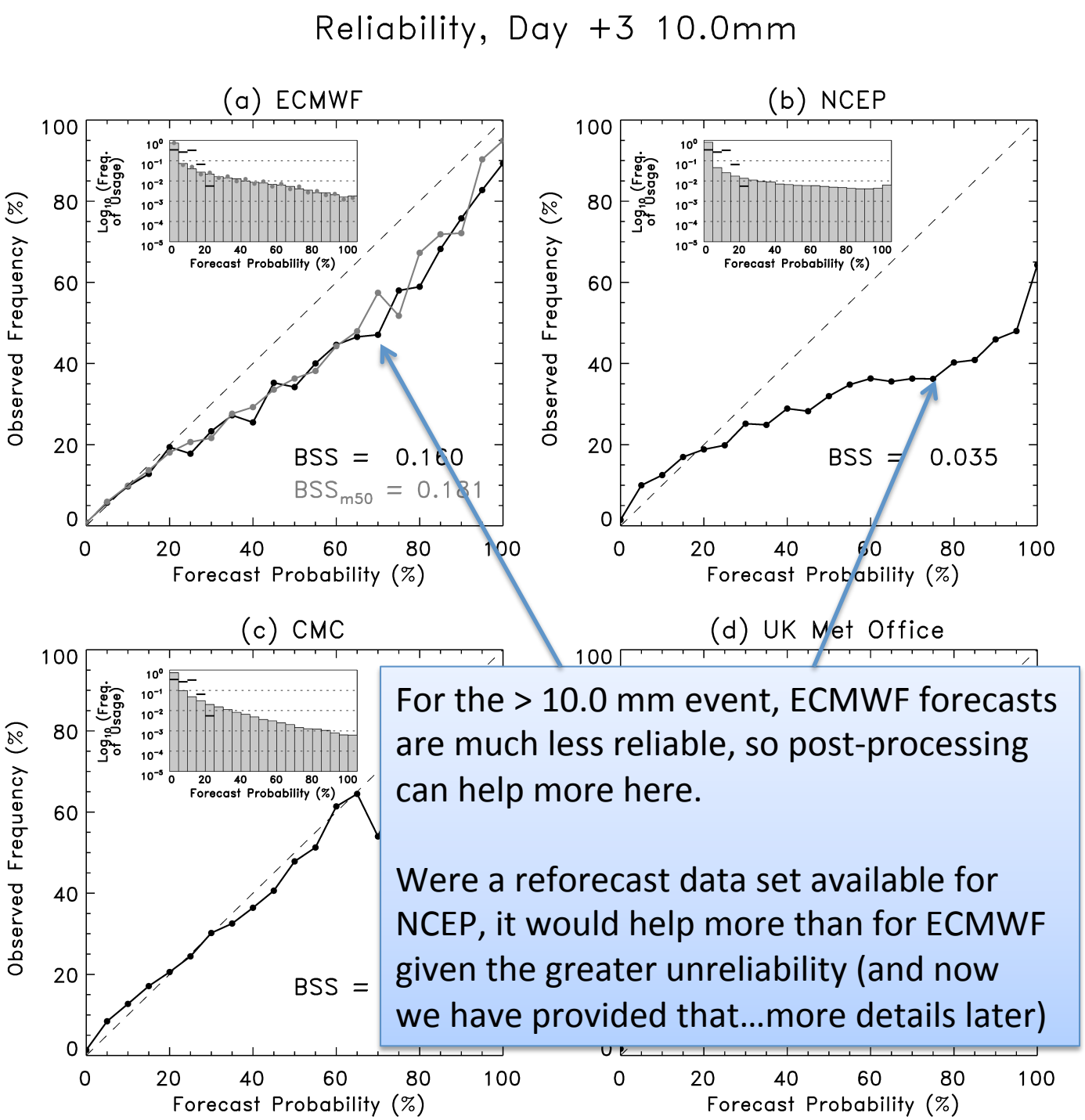
(1) Impressive skills of multi-model.
(2) Reforecast calibration doesn't improve the 1-mm forecasts much, improves the 10-mm forecasts a lot.
(3) Calibration of multi-model using prior 30 days of forecasts doesn't add much overall.



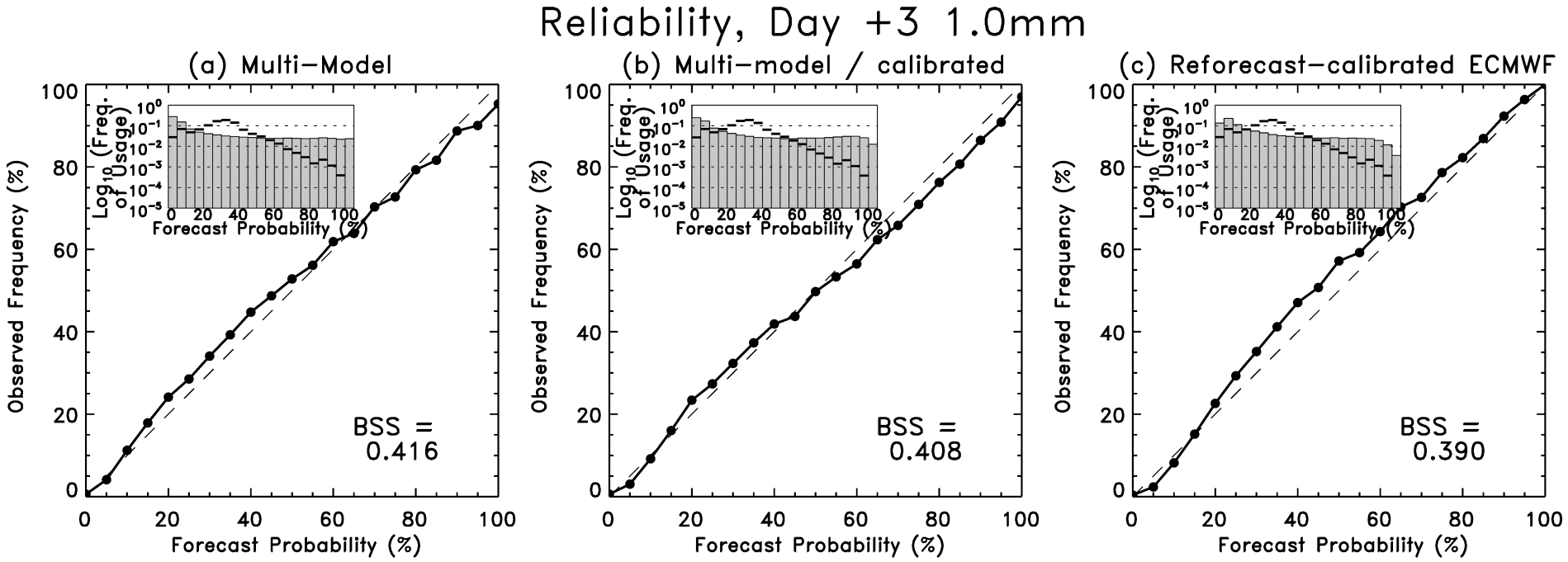
Reliability diagrams, day +3

> 1.0 mm

For the 1.0 mm event, ECMWF forecasts are relatively reliable already, so post-processing isn't likely to help as much.



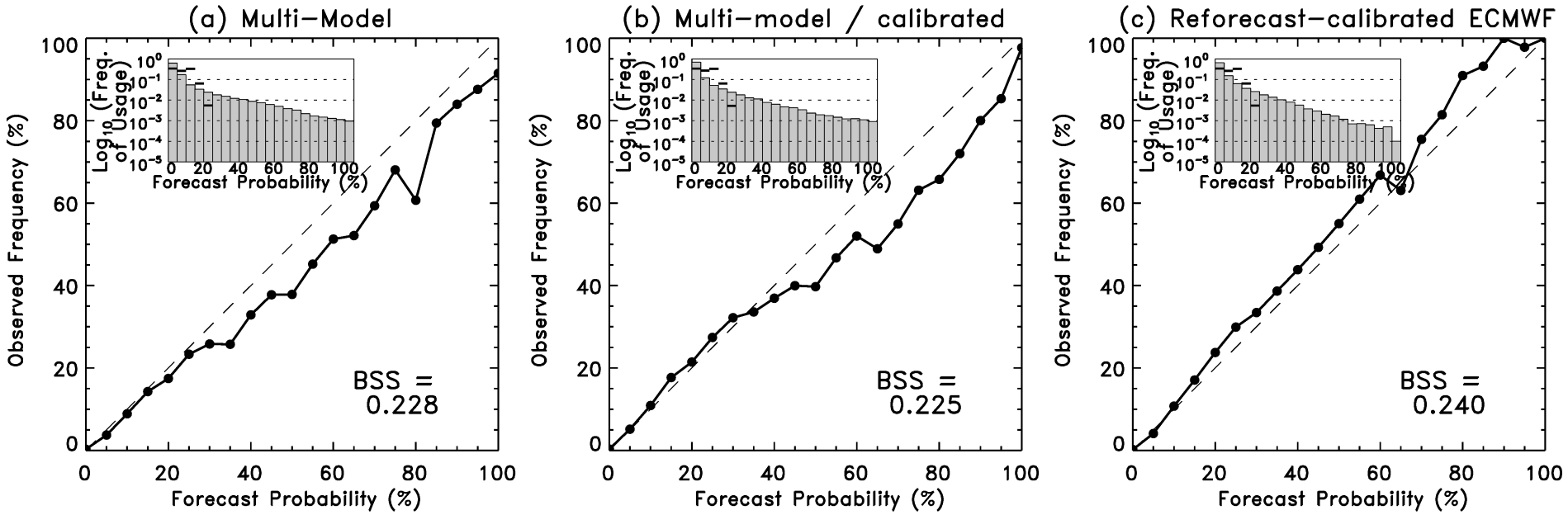
Reliability diagrams, day +3, > 10 mm



last 30 days of forecasts used to calibrate

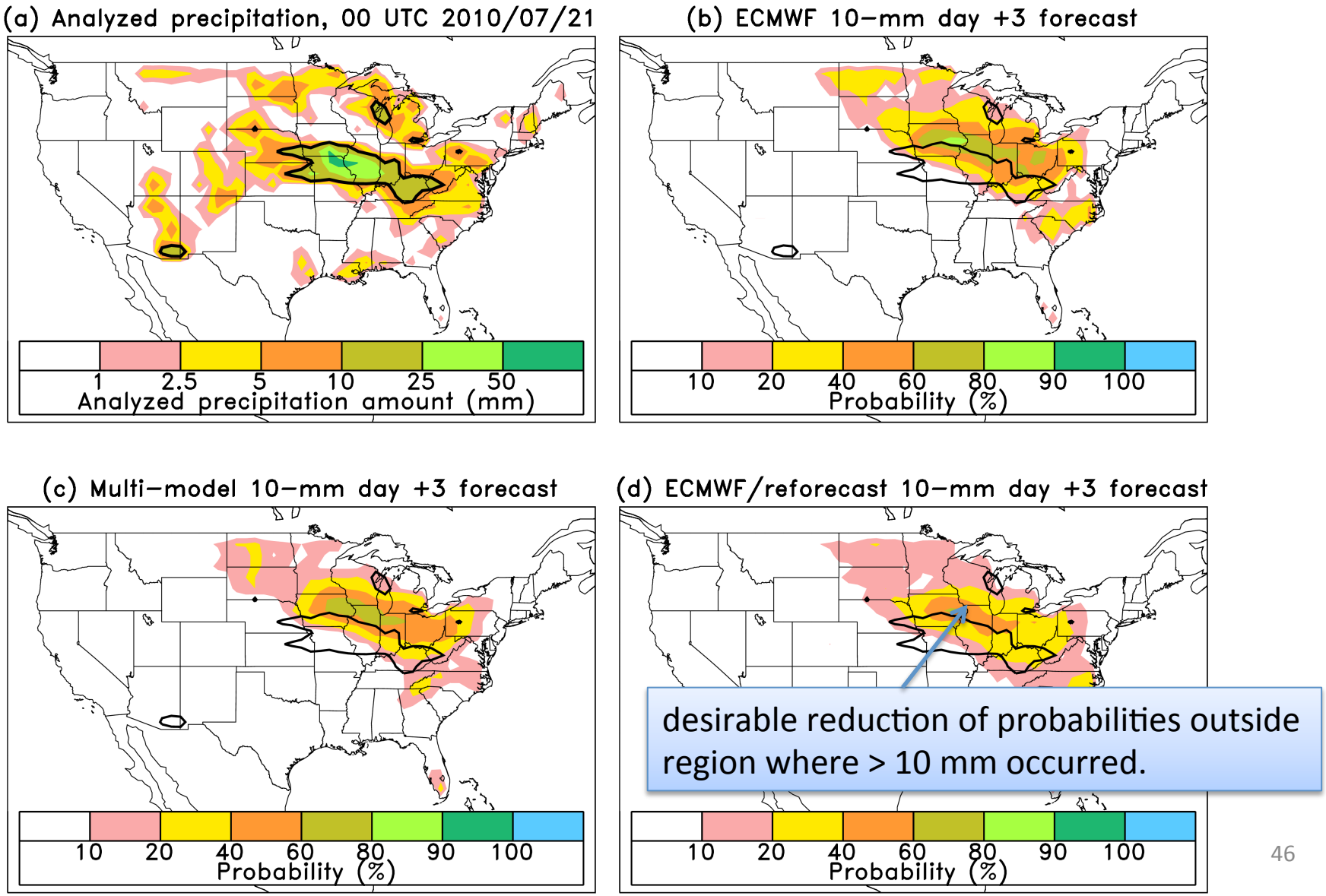
8 years of once-weekly reforecasts used to calibrate

Multi-model slightly under-forecasts probabilities at 1.0 mm and is **quite reliable.** It is also substantially sharper than reforecast-calibrated, which has slightly greater under-forecast bias.

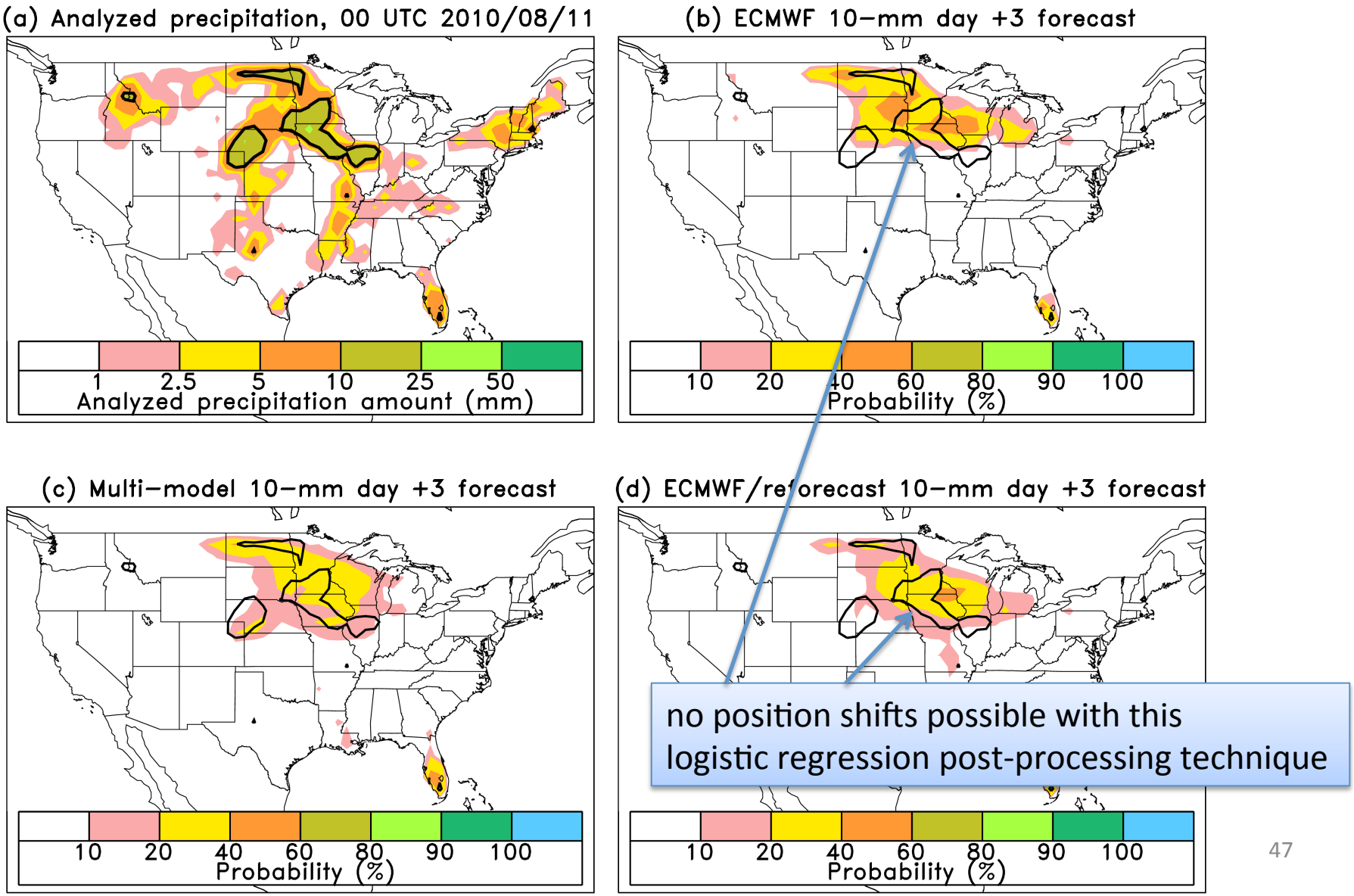


Multi-model slightly over-forecasts probabilities, and is substantially sharper. Reforecast calibrated slightly under-forecasts and is less sharp.

Forecast example: 21 July 2010



Forecast example: 11 August 2010



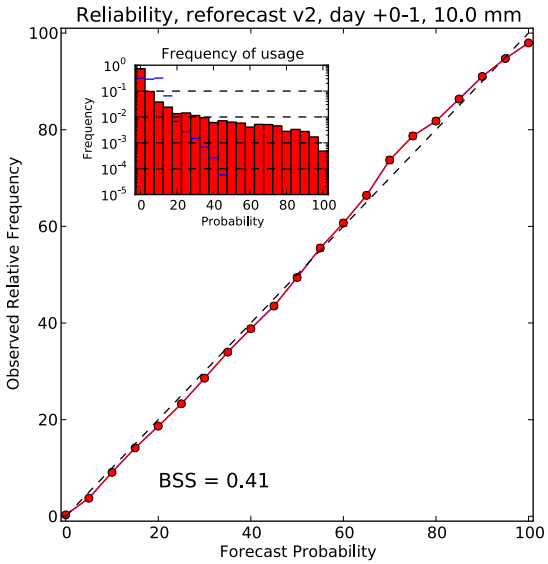
New! 2012 GEFS reforecast data set

- Developed by ESRL (on DOE computers) for 2012 NCEP GEFS.
- Every day, 1985-present, we have 11-member ensemble reforecasts computed to day + 16 using operational T254/T190 NCEP GEFS for 00Z cycle.
- CPC, EMC, HPC, MDL using this data for product development. More to follow. We hope to attract companies and universities, forecast offices to explore using it also.

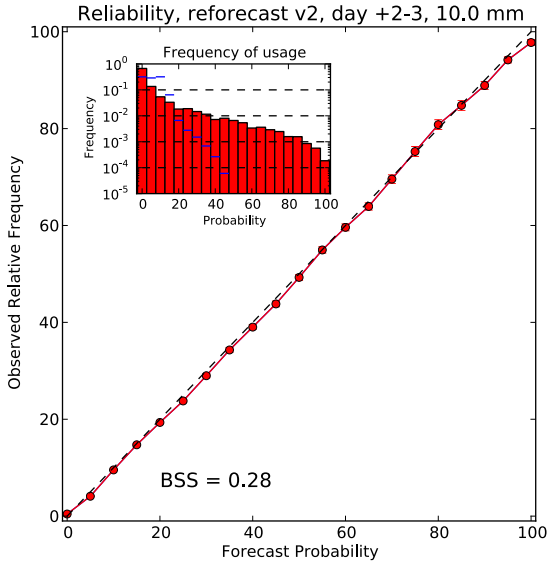
Reliability, > 10 mm precipitation 24 h^{-1}

Version 2 (2012 GEFS)

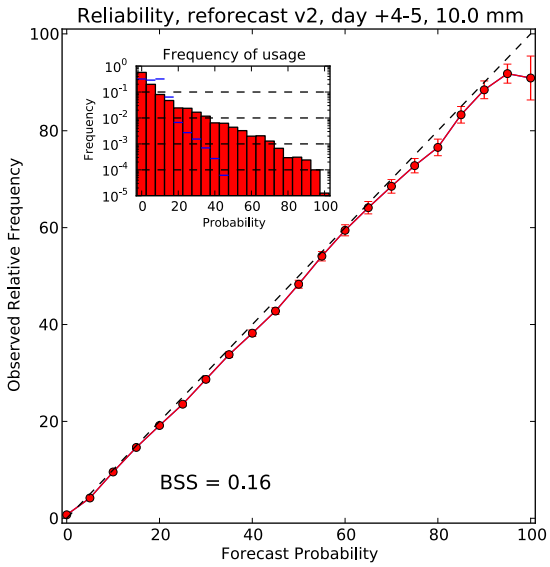
Day +0-1



Day +2-3

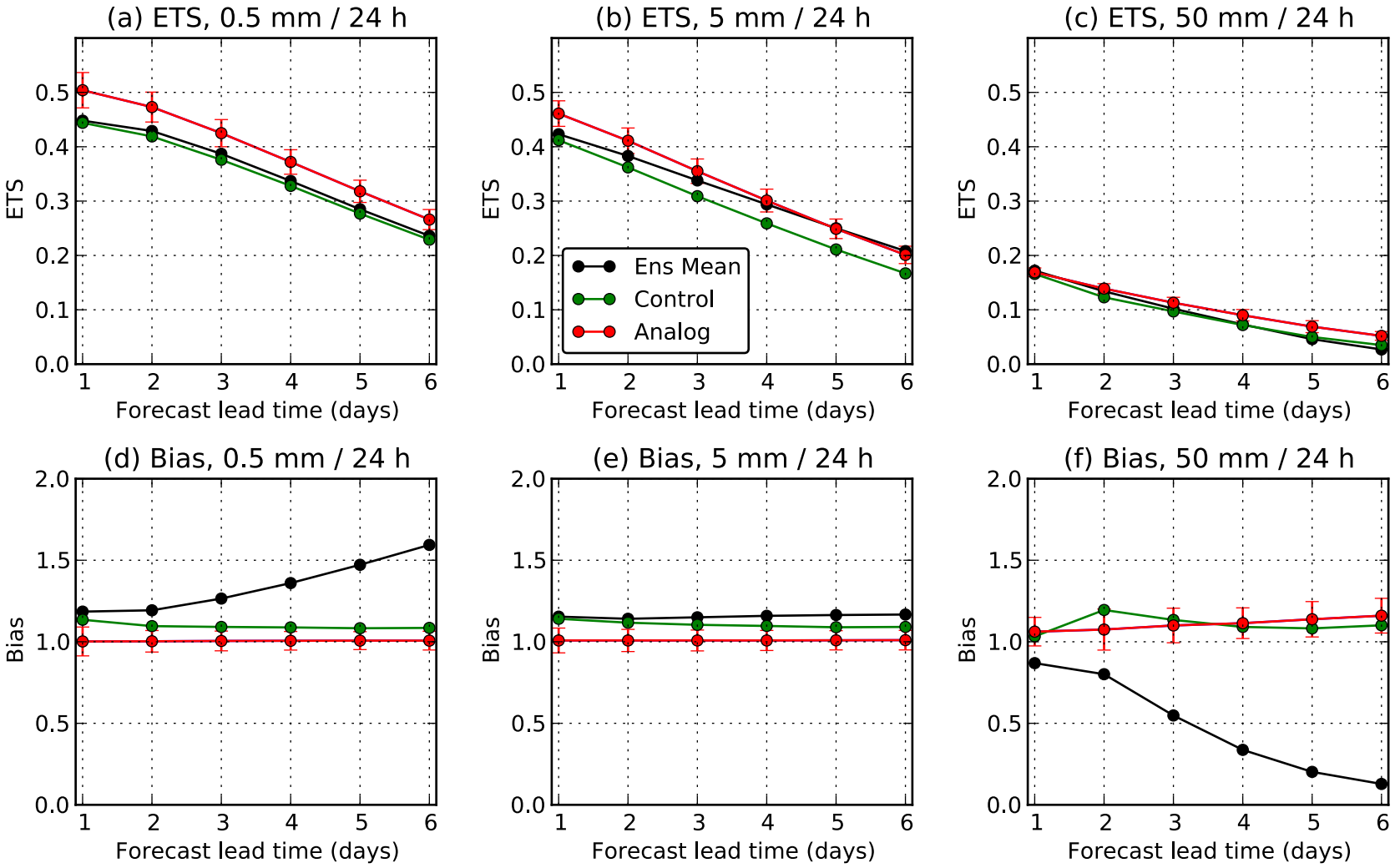


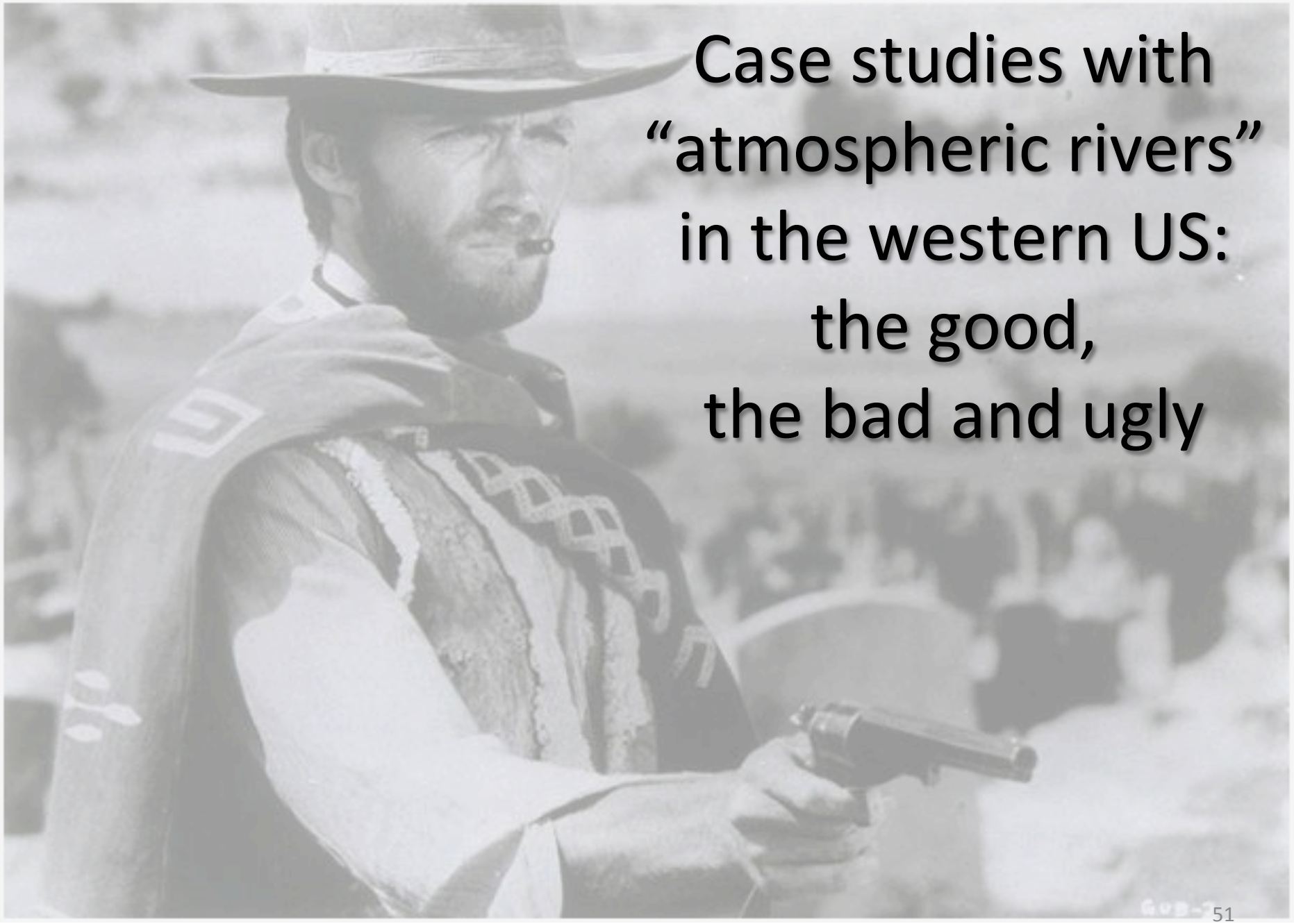
Day +4-5



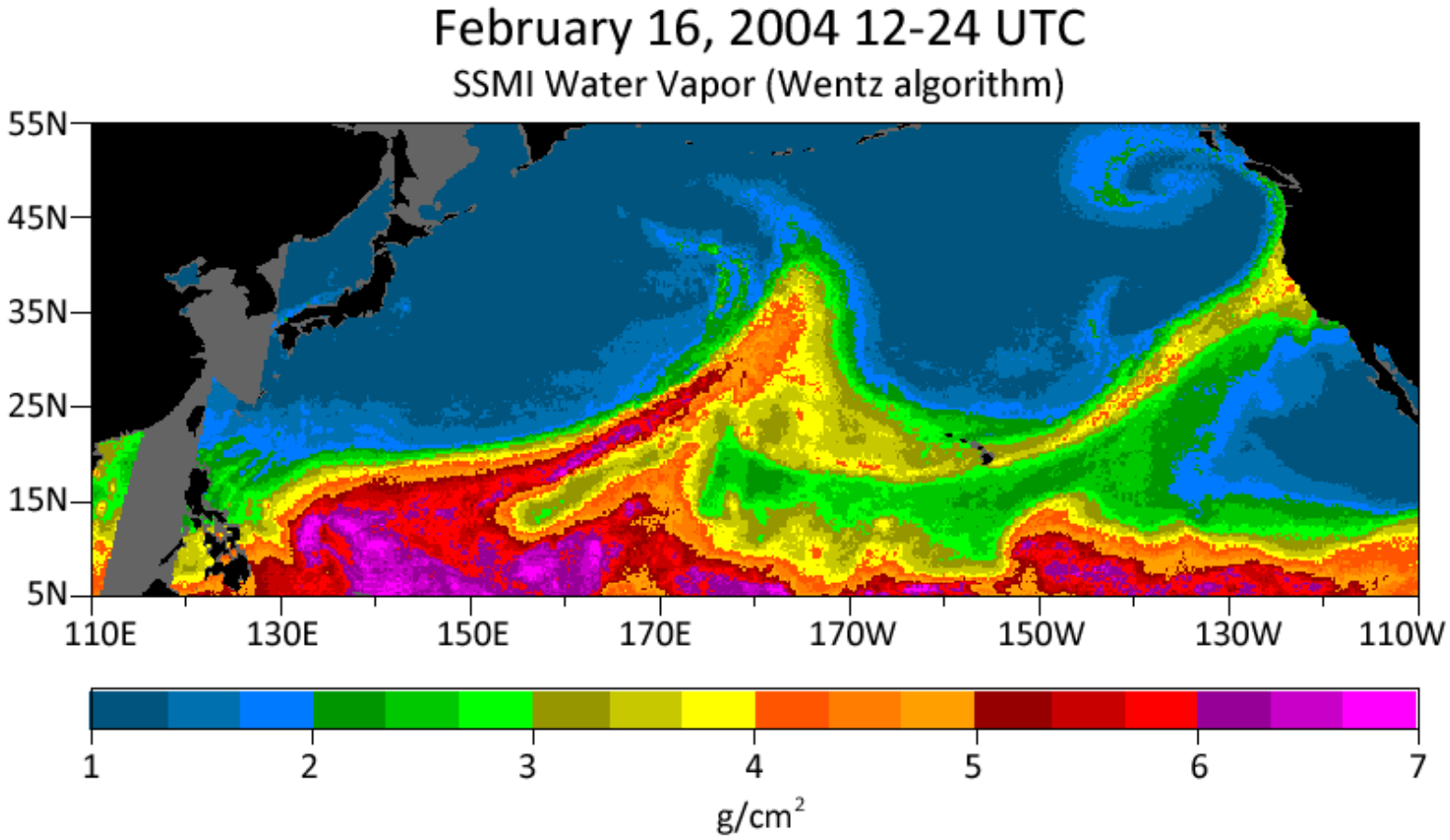
We can make very reliable (and skillful) probabilistic precipitation forecasts by post-processing the GEFS using the reforecasts.

Example: improving deterministic precipitation forecasts with statistical post-processing.



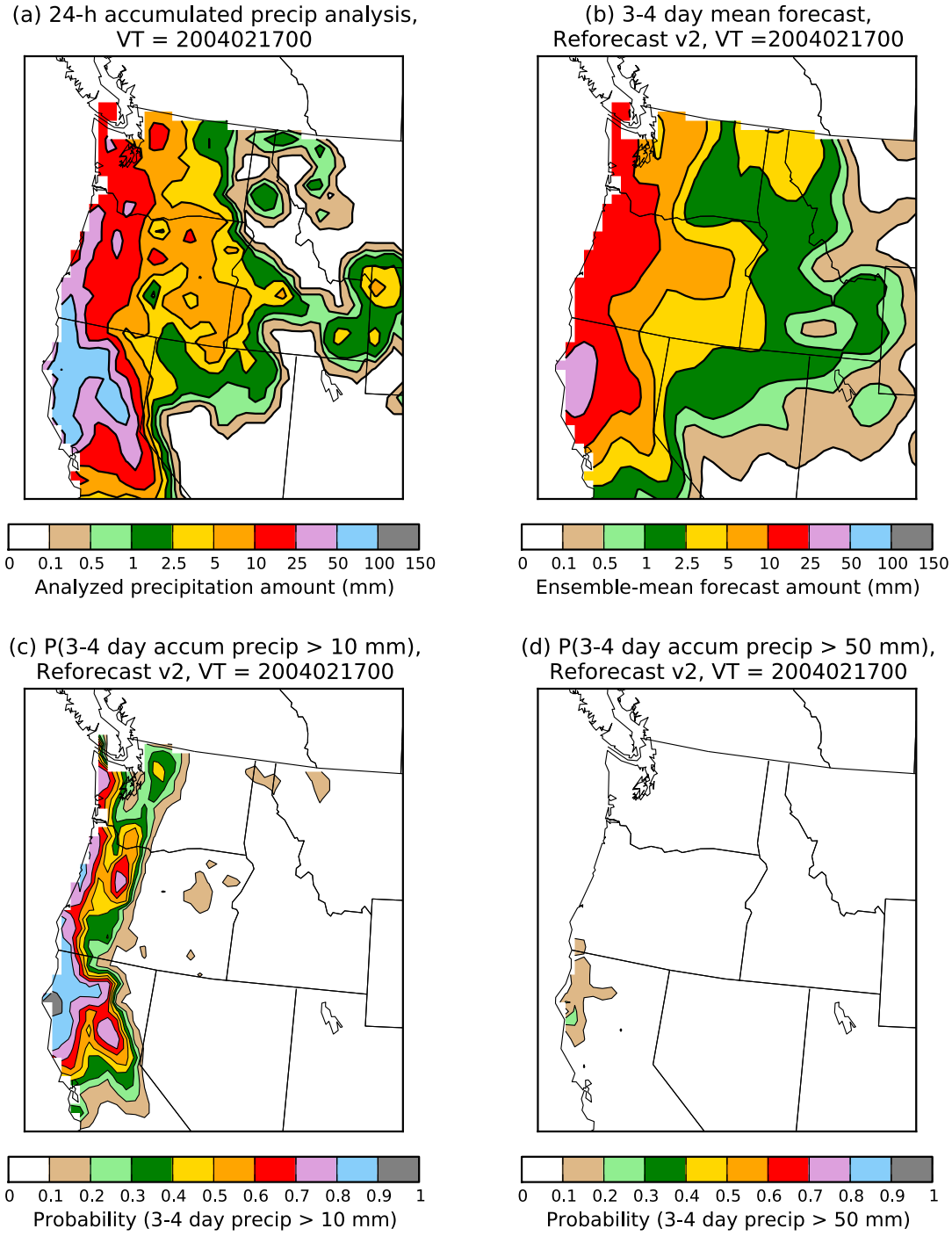

Case studies with
“atmospheric rivers”
in the western US:
the good,
the bad and ugly

The bad and ugly atmospheric rivers case study

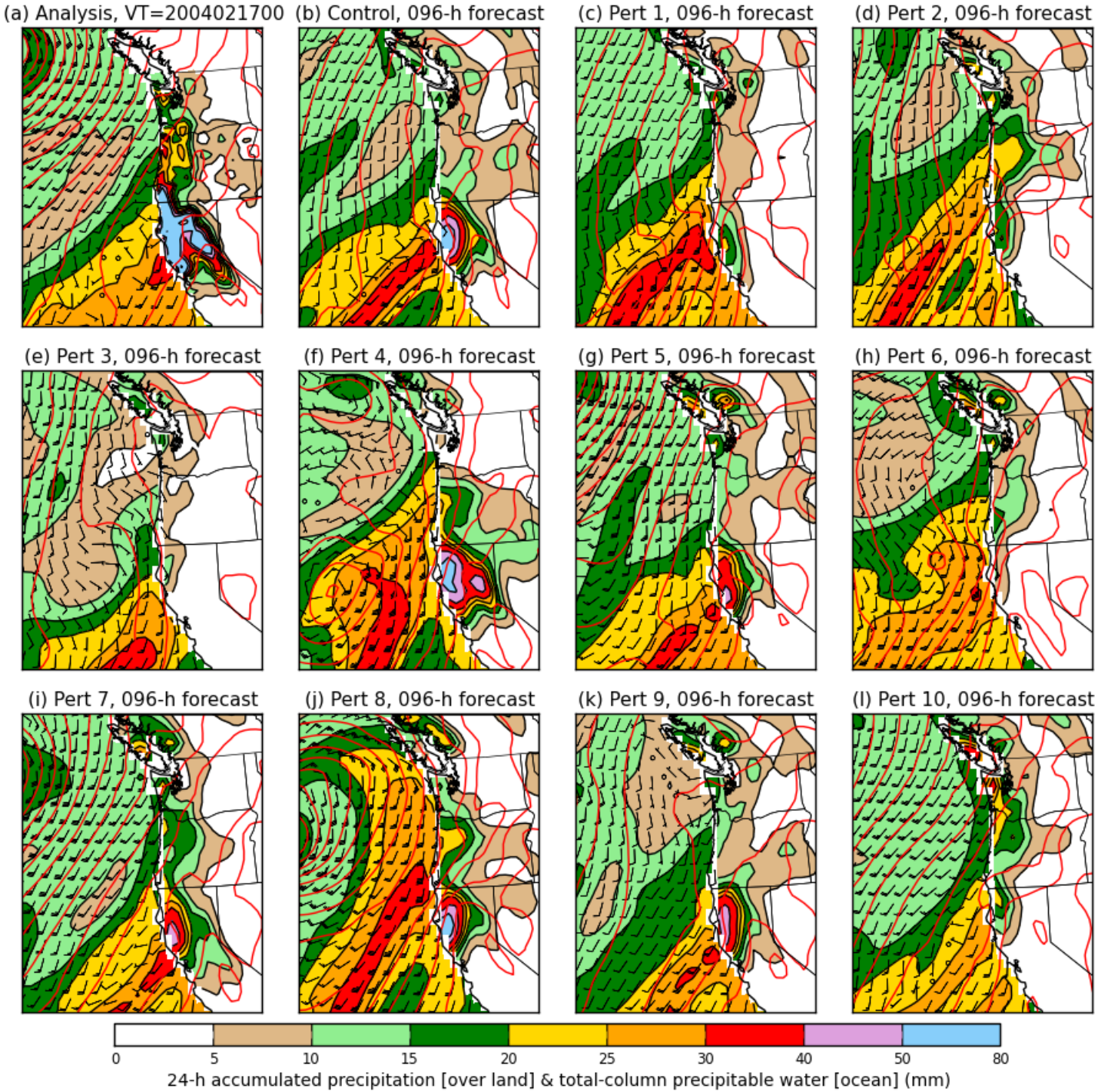


➢ 10” rain in the coastal mountains, 4-7” in Russian River watershed. Streamflows in top 0.2% of historical observations.

Source: http://www.esrl.noaa.gov/psd/atmrivers/events/



4-day forecast

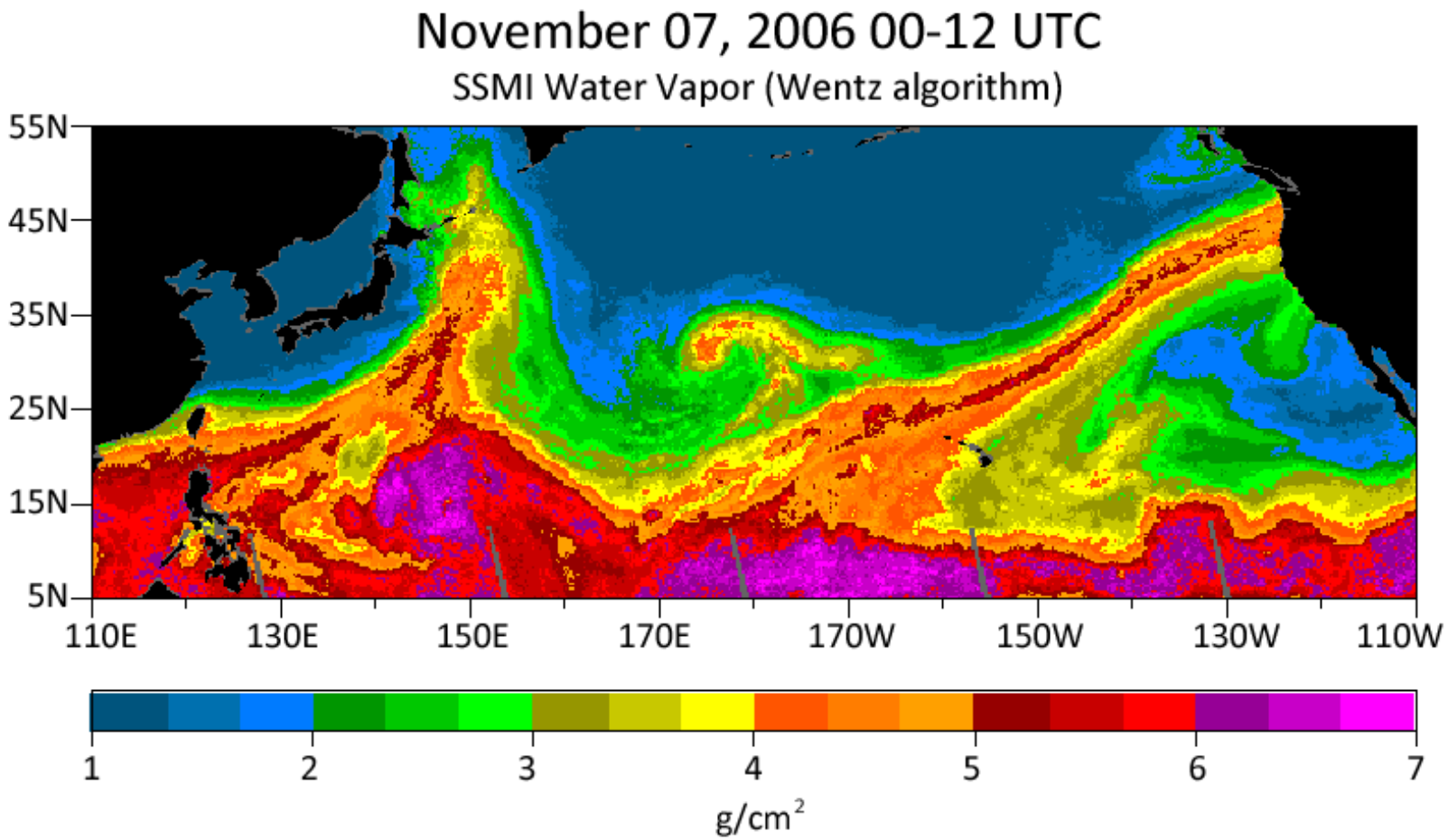


4-day forecast

Case study, tentative conclusions

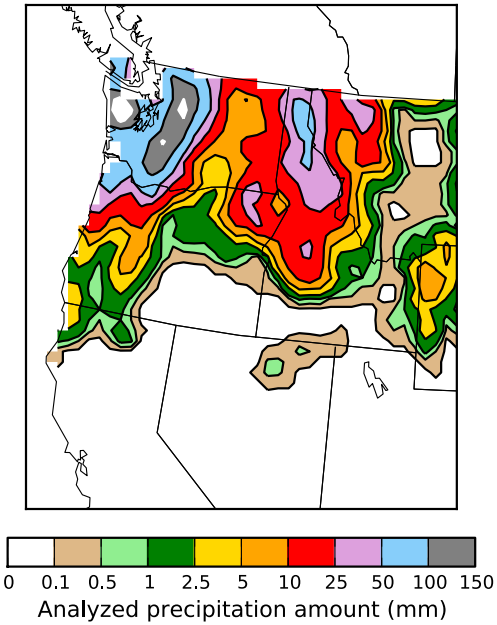
- Statistical post-processing will not be able to correct for everything. In this case, the synoptic-scale predictability was apparently quite low.
- Improvements to post-processed probabilistic forecasts in such a case will require improved ensemble guidance.

The “good” atmospheric rivers case study:
Nov 2006 Oregon-Washington floods

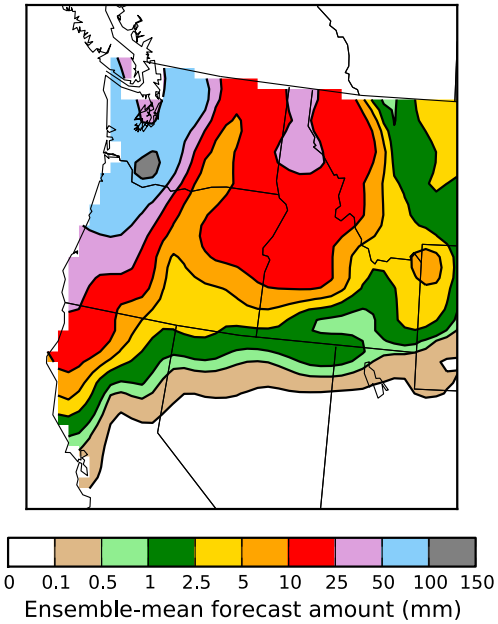


8-20 inches of rain in Cascades; flooded rivers; extensive damage to Mt. Rainier NP.

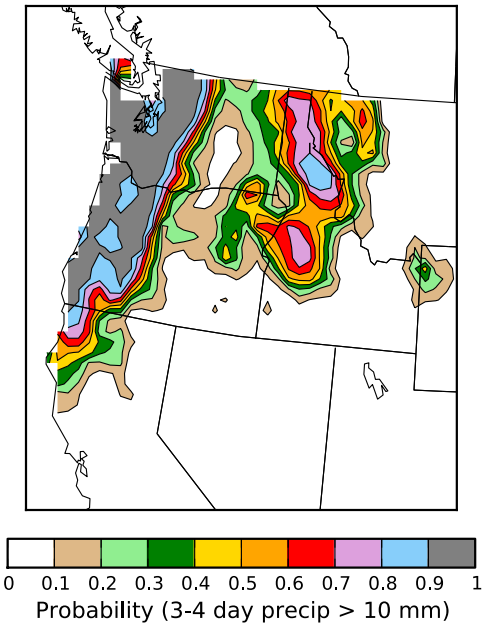
(a) 24-h accumulated precip analysis, VT = 2006110700



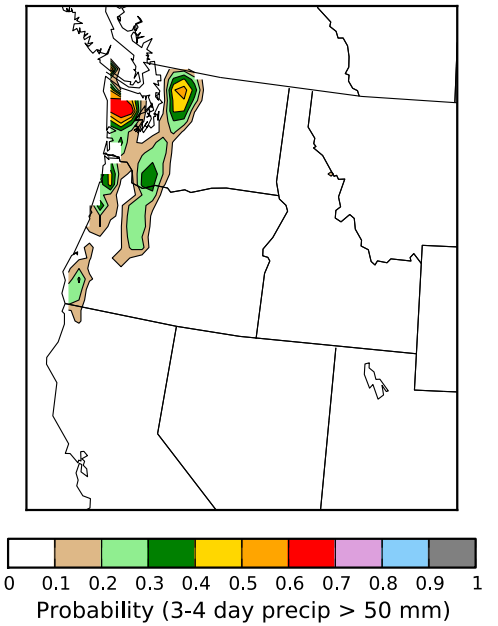
(b) 3-4 day mean forecast, Reforecast v2, VT =2006110700



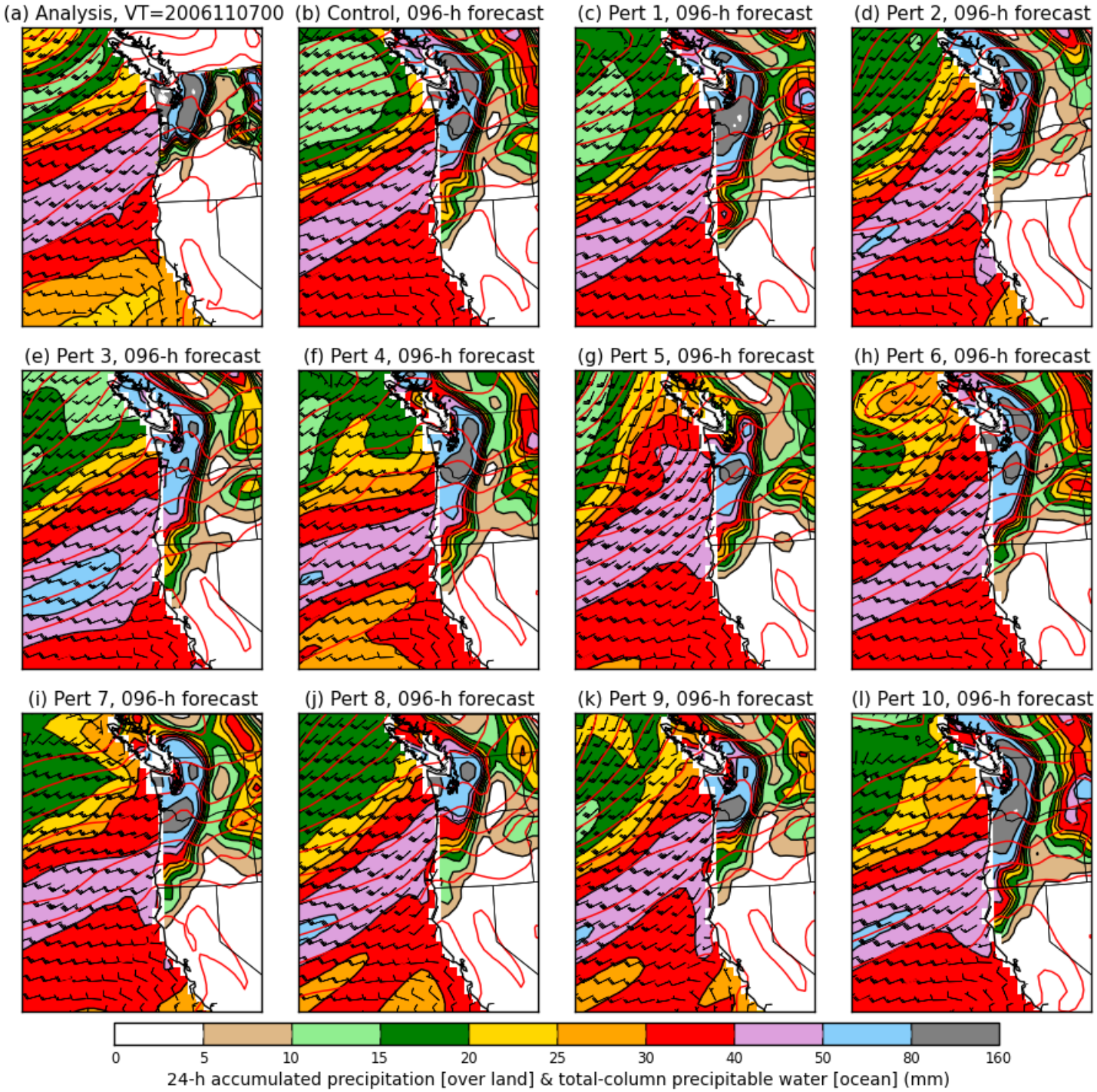
(c) P(3-4 day accum precip > 10 mm), Reforecast v2, VT = 2006110700



(d) P(3-4 day accum precip > 50 mm), Reforecast v2, VT = 2006110700



4-day forecast



4-day forecast

A few words about regional ensembles

- Benefits:
  - you can probably afford much higher resolution than you can with global ensemble, so less model error.

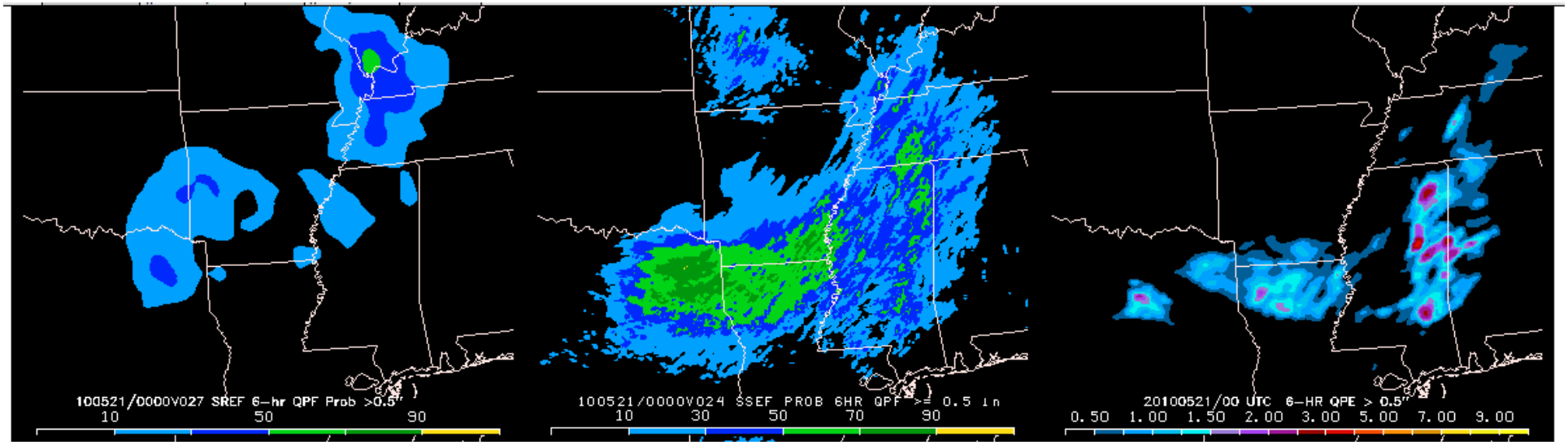
- Disadvantages: boxed in.
  - use of regional models in general causes some problems; errors can't propagate up to planetary scales; lateral boundary conditions introduce errors; may be inconsistencies between global and regional.

Regional ensembles and explicit convection

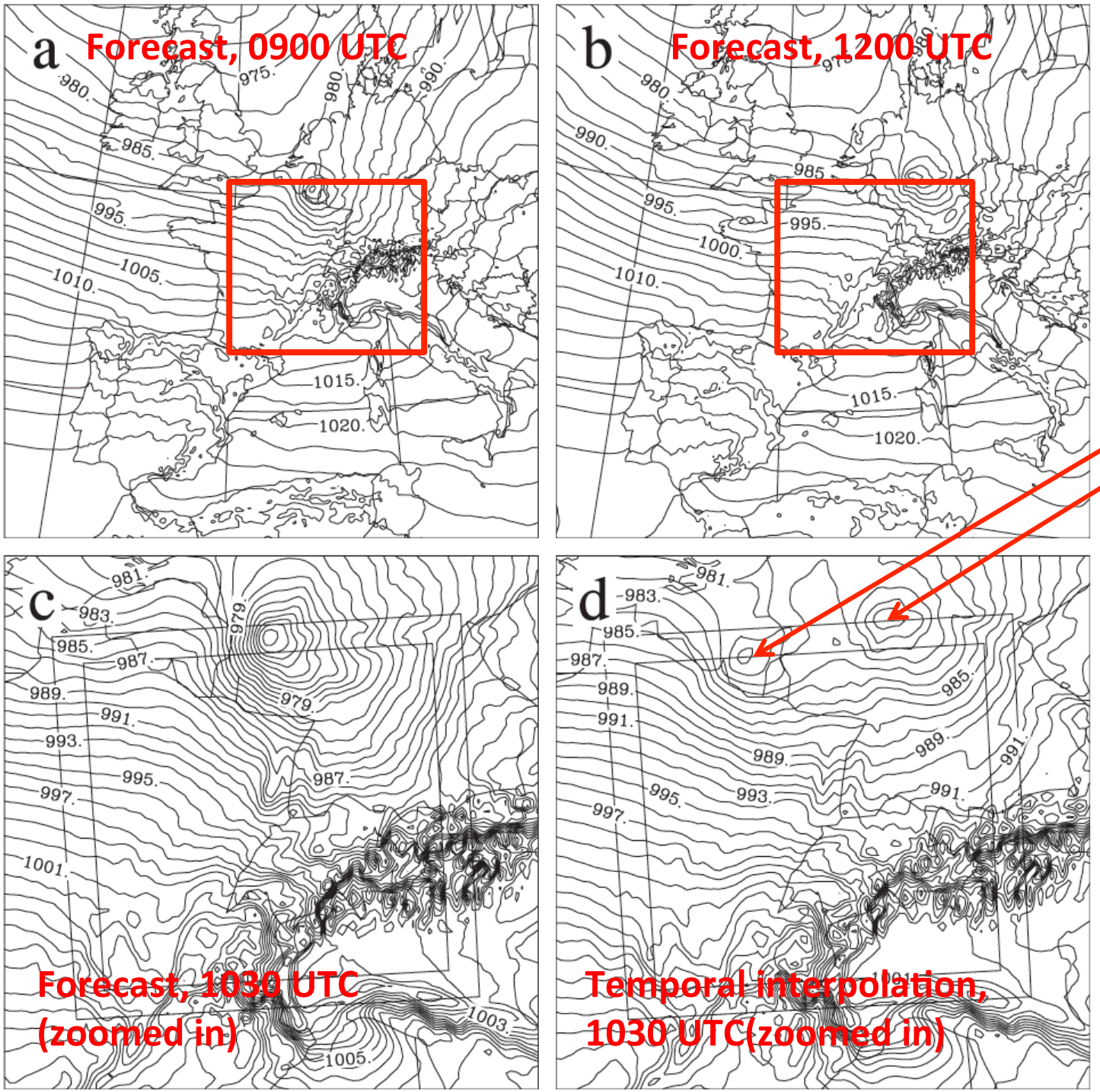
An example from NSSL-SPC Hazardous Weather Test Bed, forecast initialized 20 May 2010

http://tinyurl.com/2ftbvgs

30-km SREF P > 0.5” 4-km SSEF P > 0.5 “ Verification



With warm-season QPF, coarse resolution and parameterized convection of SREF clearly inferior to the 4-km, resolved convection in SSEF.



Example: interpolation errors in Aladin simulation of 1999 Lothar storm

temporal interpolation creates two lows from one.

Ref: Tudor & Termonia, MWR, July 2010

Conclusions (part 1)

- Producing a high-quality ensemble for you takes more than slapping a few forecasts together.
- We need to (and are) designing ensembles that address initial-condition uncertainty and model error.